THE ROLE OF SELF-HEARING IN SPEECH PRODUCTION

CAROL ANN SHERRARD

Ph.D., University of Edinburgh, 1980.

<u>DECLARATION</u>

This thesis has been composed by myself, and the work it reports is my own.

Carol Ann Sherrard

ACKNOWLEDGMENTS

I thank first and foremost my supervisors, Terry Myers and John Marshall, for their generous help and interest in this work. They often seemed to know what I meant before I did.

Thanks are also due to David Lee, Roly Lishman, Shimon Abramovici, and Brian Millar for clarifying discussions.

Members of the Psychology Department's Electronic Workshop gave valuable help in designing and setting up apparatus which did the difficult job of converting vague ideas for experiments into realities.

Judy Hutton furthered the quest for clarity by her careful typing of the thesis.

Last, but far from least, I thank my husband Arthur Still for help, not only with methodology and statistics, but with everything else that counts.

ABSTRACT

The view that speech production is under auditory feedback control is discussed and experimentally evaluated with reference to general theories of motor control and observations of speech in abnormal circumstances. In particular, an alternative view (Lane and Tranel's) is assessed which claims that changes in speech performance in different acoustic conditions reflect social, and not auditory control, and that these changes tend to preserve intelligibility rather than to impair performance.

Experiments I-V, on the control of voice level, pit these two views against each other. Subjects carried out linguistic and non-linguistic vocal tasks with varying access to auditory and social feedback. The findings were that subjects could readily make relative changes in voice level with masked self-hearing, but that a task requiring absolute voice matching and accurate judgment of one's own voice level was impossible. Further, the relative voice level changes possible without self-hearing were not responsive to social factors per se, but reflected a stereotyped emergency response to any communication difficulty, which was simply to speak louder as the difficulty increased. This response, though crude, would tend to preserve intelligibility.

The impairment of speech produced by deafness and by delayed auditory feedback (DAF) have been seen as evidence of the indispensability of sensory feedback in motor control. Shortcomings of this argument are discussed, and Experiment VI shows that while feedback changes which are contingent on performance units (as in DAF) may be disruptive, changes which are not contingent may actually improve performance. Experiment VI further tests this by demonstrating that speech produced with non-contingent auditory interference can be more intelligible even than normal speech.

It is concluded that the DAF effect is a special case due to the contingency of the altered feedback, and it thus throws little light on the natural auditory control system. A model is proposed in which voice level is under continuous closed-loop auditory control, while articulation under open-loop control. When auditory feedback is masked, however, (as in the Lombard effect), voice level comes under an emergency open-loop control system which acts to preserve intelligibility.

C O N T E N T S

Page

CHAPTER ONE

HISTORY OF IDEAS ON THE ROLE OF SELF-HEARING IN SPEECH

1. Introduction

The assumption is commonly made that speech production is monitored and controlled by self-hearing. Many authors writing generally on speech will make a passing reference to this, often treating it as an established or self-evident fact:-

> "To the normal man the sound of his voice is necessary to the proper regulation of its tone and intensity." (Kerrison, 1918, p 664)
>
> "... one is guided more or less unconsciously by the sound of the words that are spoken. When one's voice becomes too loud or too high pitched, it is brought down to the usual intensity or register. The adjustment is not done consciously, nor is one aware of the departure from the normal or usual; but the ear almost reflexly or automatically checks the loud and the high tones and brings them down to the usual and normal. The lack of this control in the deaf or the hard of hearing makes itself felt in the monotonous lack of intonation that these individuals show in their conversation." (Pillsbury & Meader, 1928, p 110)
>
> "... the distance the tongue moves, the speed at which it moves, the exact moment at which it changes direction - all these are determined largely on the basis of information supplied by the auditory feed-back." (Fry, 1957, p 444)
>
> "The hearing of one's own voice permits the control of sequential patterns of vocal sounds, as in bird songs, and the monitoring of vocal utterances, as in human speech" (Gibson, 1966, pp 75-6)

In similar vein, writers who provide block diagrams of the 'communication process' will usually include an ear-voice, or ear-brain link within a speaker (Morton, 1970). Denes and Pinson (1963) include such a link in their well-known SPEECH CHAIN, but deal with it in much less detail than the other links in the chain.

That speech is controlled by self-hearing is indeed considered so certain that this supposed control system is frequently cited as the prime example of a skill controlled by feedback:-

> "It can be shown experimentally that any interference with ... feedback is liable to have drastic effects on performance. For example, it is very difficult to control the volume of one's voice when one can no longer hear oneself speak." (Beloff, 1973, p 169)

> "Most biological/psychological examples of feedback loops are negative, with more or less adequate transformation rules. An example of a nonoptimal transformation rule is found in the feedback system controlling the production of a steady note by a trained singer. ... The voice fluctuates between a note of too high a pitch and one of too low a pitch. This occurs because the comparator or test phase cannot make a fine discrimination of when zero discrepancy has occurred. If the test phase has been controlling the operator in such a way as to reduce a discrepancy (by raising the tone, for example), it will continue to do so until it begins to detect a discrepancy in the opposite direction. If the comparator is not very sensitive this change of direction (to lower the tone, for example) will not occur until the discrepancy is quite marked. In such a system the operator is used to make behaviour (voice production) 'hunt' to and fro between certain limits. In singing this can sometimes be heard as a tremolo in the voice." (Annett, Morris, Holloway & Roth, 1974, p 34)

This evidently well-entrenched view of the role of self-hearing in speech has, as we shall see, been in existence for a long time, but it was not until Lombard reported his findings (1909, 1911) that there was any experimental support for the view. Lombard, a French otolaryngologist, demonstrated that the application of loud noise to one or both ears of a speaker results in an increase of vocal intensity - henceforth known as the Lombard Effect.

This effect does not, however, necessarily constitute evidence that speech is under the control of self-hearing. We will later consider in detail the status of this and other similar 'auditory feedback' effects as evidence for auditory control of speech. For the moment, it is sufficient for the argument to indicate that Lane and Tranel (1971) have extensively reviewed the literature on 'auditory feedback' effects, and pointed out that, since the effects are more marked when there is a listener present with the speaker, they could just as well be due to the speaker's efforts to remain intelligible against auditory interference, as to perceptual-motor skill disruption through manipulations of 'feedback'.

2. Pre-Modern Belief in the Role of Self-Hearing

It is of some interest, then, to ask what were the reasons for the belief that speech is controlled by self-hearing before Lombard's demonstration, and are they still acceptable reasons?

There seem to have been four sources of evidence in favour of the belief. These were: observation of the poor or absent speech of the congenitally deaf, observation of the deterioration of speech following adventitious deafness, observation of the effects on voice of temporary hearing deficits, and finally the general notion of the control of movement by sensory feedback.

These are four possible sources of evidence that I have been able to locate, although, apart from the fourth, the general notion of sensory control of movement, there are few reasoned arguments based explicitly on them. Nor can 'common sense' be appealed to as the basis of the belief, for, although the question "Why are people who are born deaf usually unable to talk?" is included in a modern intelligence test (Wechsler, 1955, p 10), with full marks awarded only for " a knowledge that one must hear sounds or speech before being able to repeat them" (p 57), it was formerly thought that in deaf-mutism either the organs of speech were coincidentally paralysed (Nottingham, 1857, refers to this belief) or that a defect in either the ear or the speech organs would affect both, through the physical connection between them. Thus Aristotle had observed that irritation of the external meatus of the ear could bring about a cough, and had reasoned from this that there is an intimate connection between the ear and the lungs, which would explain why dumbness accompanied deafness. Nottingham (1857) also refers to an observation by Galen "that opium put into the ear, for the purpose of relieving pain of this organ, has often caused mutism" (p 471) and Pritchard had to remind his readers in 1886 that mutism accompanying deafness "is not the result of organic change, but simply a natural consequence of the person's inability to imitate the voice of others and to appreciate the results of his own efforts of articulation" (p 173). In fact, this physical or mechanistic view of the connection between hearing and speech continues in the present, among the practitioners of what I have called the medical tradition.*

Clearly, then, the belief that speech is under the control of self-hearing has not always been a common-sense one, if only because the more general link between speech and hearing, itself, has not always been seen for what we now know it to be - primarily a psychological link through which the learning of speech takes place.

We will now consider in detail the four sources of evidence, which existed before Lombard's demonstration, for the belief that speech is controlled by self-hearing.

(i) Deafness and Speech

The congenitally deaf

In 1860, Toynbee wrote that:

---

* See p 16ff

> "... the extreme harshness and monotony of the sound produced by deaf-mutes arise from the impossibility of regulating the tones of a voice which they cannot distinguish." (p 412)

Cassells (1883) also wrote about deaf-mutes:

> "Some deaf-mutes learn to speak so distinctly that they can without difficulty be understood by everyone ... (but) ... even in the case of those who can be easily understood there is an unpleasant harshness of utterance, as the deaf-mute has no guide to the modulation of his voice." (p 773)

Here we have the clear implication that the poor speech of the congenitally deaf is due to lack of self-hearing, rather than to failure to hear the models of the sounds to be produced in the first place, a somewhat odd emphasis which persists to the present day, and which testifies to the strength of belief in the role of self-hearing.

The adventitiously deaf

John Kitto went totally deaf at the age of twelve, as the result of a fall from a ladder. In his book THE LOST SENSES (1860) he quotes from a friend's description of his speech:

> "It is pitched in a far deeper bass tone than is natural to men who have their hearing. There is in it a certain contraction of the throat, analogous to wheezing; and altogether, it is eminently guttural. It may be suspected, that this is attributable to the fact that his deafness came on in boyhood, before the voice had assumed its masculine depth. The transition having taken place without the guidance of the ear, was made at random ... His pronunciation is generally accurate enough, as regards all such words as young boys are likely to be familiar with, and as to others which closely follow their analogy: but it is naturally defective in respect to words of later acquirement." (DEAFNESS, p 22)

Kitto also comments on his loss of ability to judge the projection of his voice over a distance:

> "... I never recovered the power of pitching my voice to a distance, or the confidence of being able to make myself understood by a person not close at hand; although I am told that my vocal powers are not in this respect so deficient as my own impressions seem to intimate." (DEAFNESS, p 141)

Note that in both these quotations, the role of self-hearing in normal speech is taken entirely for granted. Similarly, Roosa (1873) writes:

> "Persons who become completely deaf later in life, do not lose the power of speech; but they usually speak in an unnatural tone, because they are unable to hear their own voice with distinctness." (p 515)

<u>Temporary hearing deficits</u>

In addition to the effects of congenital and adventitious deafness on speech, it had been observed before Lombard's demonstration that a temporary deficit in hearing could affect the ability to judge one's own voice production. Brown and Behnke (1884) write that, if the Eustachian tubes are closed by swelling, as during a cold or sore throat,

> "... then the hearing is impaired, and speakers and singers are particularly inconvenienced by not being able accurately to judge the sound of their own voices ... One striking case of this nature occurred to us in which a very well-known actor, formerly under our care, came later blaming us for not having recognised some serious diseased condition of his larynx ... But he was not aware until his hearing was tested that the auditory function had simultaneously deteriorated. On syringing the ears very large plugs of impacted wax were removed ... with the immediate result of complete and permanent vocal improvement." (p 155)

Mackenzie (1893) evidently entirely accepts this viewpoint:

> "The ear is the conscience of the voice, and its purity should be not less jealously guarded. Many singers of the finest vocal endowments fail from a defect of ear; their condition is like that of a colour-blind painter. Passing indisposition may sometimes vitiate the ear as well as the temper; the artist should on no account attempt to sing under such circumstances." (p 100)

(ii) <u>The general model of the sensory inflow control of movement</u>

As we have found few reasoned arguments so far in favour of the model of speech control by self-hearing, it seems likely that the notion was based either entirely on informal observation, or, what is perhaps more likely, on a prevalent belief that movement is controlled by sensory inflow, a belief which could have implicitly influenced the motivation and <u>interpretation</u> of the sort of observation we have been quoting.

A popularising physiology book of the period by Lewes (1860) suggests that this is the case. He clearly puts forward the view that movement is under sensory control, at least in the acquisition phase:

> "We have acquired a power of definite direction in the movements of the hands, which renders them obedient to our will, but this acquisition has been of slow laborious growth. If we were asked to use our toes as we use our fingers - to grasp, paint, sew, or write with them, we should find it not less impossible to control the movements of the toes in these directions, than to contract the iris, or cause a burst of perspiration to break forth. Certain movements of the toes are possible to us; but unless the loss of our fingers has made it necessary that we should use our toes in complicated and slowly acquired

.. ..

> movements, we can do no more with them than the young infant can do with his fingers. Yet men and women have written, sewed, and painted with their toes. All that is required is that certain links should be established between sensations and movements." (p 155)

Interestingly enough, Lewes the physiologist cites the psychologist Bain (1859) as the authority for this view. Bain, however, cites no authority in the referenced work.

The confidence with which such views were put forward suggests that, even allowing for the relatively unscholarly writing of the time, no authority was in fact needed. The idea that movement is under sensory feedback control was, and had been for a long time, in the air. Vartanian (1960) traces the origin of psychological cybernetics to La Mettrie, a contemporary of Descartes who took Descartes' dualism to its logical conclusion by asserting that, since the body was able to operate (mechanically, at least) in complete independence from the soul, the soul was redundant in descriptions of behaviour. La Mettrie thereby founded modern mechanistic psychology.

It is also true, according to Vartanian, that mechanistic conceptions of psychology were revitalised from around the beginning of the nineteenth century, and it seems highly probable that this was due to the occurrence in 1807 of what Brett (1912) has called "the most important event in the early history of neurology". This event was Charles Bell's separation of the functions of the posterior and anterior spinal nerves into sensory and motor respectively, and his proposal that information from the sensory nerves was relayed back via the motor nerves to control movement. The finding was first published in 1881, reported to the Royal Society in 1821, and reported more fully in Bell's NERVOUS SYSTEM in 1830.

The discovery was confirmed by Magendie in 1882, and established more explicitly by him as the distinction between efferent and afferent systems (Brett, 1912).

It is easy to see that these physiological discoveries would set a seal of scientific respectability on a pre-existing notion whose origin had probably been forgotten. Thus Boring (1942), Ruch (1951), and Miller, Galanter and Pribram (1960) all trace to Bell's 'circle of nerves' the general notion that sensory information is sent from the muscles to the central nervous system, and that this information is used in the control of movement. Chase et al (1961) have suggested that Bell's 'nervous circle' is the origin of the notion of the auditory feedback control of speech.

In fact, it is only after the Bell-Magendie discovery that we find reasoned accounts of the notion of auditory control of speech, explicitly based on theories of sensory control of movement. Thus, in 1860, Fournié references a passage in a text-book by Claude Bernard, which asserts that the section of the (posterior) sensory spinal nerve leading from a limb will deprive it of co-ordinated movement, and Fournié generalises from this to the case of speech:

> "Si, par exemple, nous considérons les mouvements compliqués de la parole, nous constatons qu'ils ne sont possibles qu'à la condition expresse que l'ouïe préside à leur formation." (pp 628-9)*

However, a quotation from later in Fournié's book reveals that he considers self-hearing to be necessary only during the acquisition of speech, and indeed he makes the interesting statement that adventitious deafness will make no difference to speech performance, other than the ability to learn new pronunciations:

* "If, for example, we consider the complicated movements of speech, we can see that they are only possible if their formation is guided by hearing."

"Dès que l'éducation de la parole est terminée ... l'ouië ... peut faire défaut; l'homme peut devenir sourd, et continuer cependant ses relations verbales avec ses semblables; mais il n'apprendra plus que très-difficilement de nouvelles dénominations." (p 646)*

By the time that Fournié was writing, or at the latest soon afterwards, Claude Bernard's concept of (as it is now known) 'homeostasis', ie of the higher organism as a self-regulating mechanism, was also well known, although this was a biochemical rather than a sensory-motor concept.

A further strand to the argument for auditory control of speech from general sensory motor control was added when Sherrington and Mott (1895) made their classic demonstration that, in monkeys, deafferentation of a limb results in loss of purposive movement of the limb. It was natural, then, for Mott to argue in his book THE BRAIN AND THE VOICE IN SPEECH AND SONG fifteen years later, at a time when there were thought to be no proprioceptors in the larynx, as follows:

"It is remarkable that there are hardly any sensory nerve endings in the vocal cords and muscles of the larynx, consequently it is not surprising to find that the ear is the guiding sense for correct modulation of the loudness and pitch of the speaking as well as the singing voice." (Mott, 1910, p 39)

(This is an argument which has appeared again as recently as 1971 (Campbell) - "proprioceptors (are) absent from the important laryngeal muscles, and feedback control comes by way of the ear" (p 14) although nerve endings and neuromuscular spindles, indicating either reflex or

---

* "As soon as the learning of speech is completed, hearing ... can be dispensed with; a man can become deaf, yet still communicate with others; but it is only with great difficulty that he will learn new words."

proprioceptive sensory regulation were discovered in the laryngeal muscles in the late 1950's and early 1960's (Lucas Keene, 1957; Rudolph, 1961) and proprioceptive receptors were located in the intrinsic laryngeal muscles by Baken in 1969.)

Mott extended the domain of auditory feedback control to articulation, as well as voice, since

> "The part of the brain concerned with the sense of hearing develops earlier and the nerve fibres found in this situation are myelinated at an earlier period of development of the brain than the portion connected with the sense of movement of the muscles of articulation." (pp 82-3)

In other words, according to Mott, the hearing parts of the brain are myelinated earlier than those receiving kinaesthetic information.

Unlike many previous writers, Mott indicates that the role of auditory control continues <u>after</u> the phase of acquisition of speech - "a child up to the fifth or sixth year in full possession of speech will become dumb if it loses the sense of hearing from middle-ear disease, unless it be educated later by lip language". (p 84)

In between Mott and Sherrington's demonstration, and Mott's sure assertion about the role of self-hearing in speech, we find Scripture, writing in 1902, using the phrase 'sensory-motor control' in relation to speech. In the light of the use that would later be made of Norbert Wiener's CYBERNETICS in explaining the delayed auditory feedback effect, it is interesting to find Scripture using the earlier 'control system' analogy of a steam-engine governor to explain voice control:

> "The SENSORY-MOTOR CONTROL is generally muscular and auditory. The action of the vocal muscles occurs under guidance of the sensations of movement obtained from them. The association of the correct movement-sensation ordinarily occurs with the aid of

.. ..

hearing the sounds produced. In the deaf it occurs without this aid ... The amount of stimulation sent to the muscles at each movement is governed by the sensations; this is followed by a reduction in the amount of nerve impulse. The reduction is generally too great; the sensations then vary in the reverse direction; and renewed correction is attempted. For a contraction intended to be constant, as of the cricothyroid in singing a tone of constant pitch, the continually fluctuating and erroneously changing motor impulses produce changes in the sensations from the tendons and in the pitch of the tone heard. (This last is not a factor of control in the deaf.) The intention to keep a constant pitch results in an adjustment of the vocal centers to receive constant sensations and to import motor impulses standing in definite relations to them. The fluctuating sensations actually received are used to regulate the impulses. An analogy may be drawn to an engine with its governor; too great speed causes the governor to reduce the steam supply, and conversely; without a fly-wheel to make the changes slow the engine would require rapid readjustments by the governor. The vocal mechanism is light and delicate; its small inertia renders its action very fluctuating; it thus requires continual regulative action. When a rising tone is desired, the governing center is adjusted so that each degree of intensity of the sensations is answered by an increased motor impulse. Falling tones are regulated by the opposite relation. A rise or fall that seems steady to the ear requires a complicated - probably not proportional, perhaps logarithmic - relation." (Scripture, 1902, pp 387-9)

3. Modern Views For and Against the Role of Self-Hearing in Speech

(i) Views in Favour

Lombard clearly qualifies as the first modern (by which we mean 'experimental' - Mott's work was speculative) proponent of the role of self-hearing in speech production, and his work, together with the cybernetic tradition, appears to be the silent underpinning of most of the writing done on this subject up to 1950, when the cybernetic rationale was made explicit. The 'silent underpinning', since very little experimental work was done during this period, and most of it consists of atheoretical technical research reports, mostly military and inaccessible (eg Fletcher, Raff and Parmley, 1918). Apart from these, the general psychological literature (eg Smith and Guthrie, 1921) and the literature on deafness (Kerridge, 1935; Carhart, 1947) continue to assume the regulative role of self-hearing, and to fail to mention any empirical grounds for the belief. (Thus one is surprised to find Piaget, in 1948, writing "the normal child regulate(s) his own phonation primarily according to the acoustic effect he notices" (1953, p 77) until one comes across Rotman's (1978) comment that "Piaget's model of the thinking mind ... can be placed in a clearly identifiable tradition. A tradition that starts from Claude Bernard's discovery of self-regulative cycles governing bodily activity, passes through Cannon's formulation of homeostasis, then the cybernetic models of intelligence produced in the early 1940's, and rests now in the computer-dominated discipline of Artificial Intelligence." (p 373).)

1950 was the year in which Black appealed to a "feedback principle" governing speech production, and in which Lee first reported the delayed auditory feedback (DAF) effect, whereby returning a speaker's recorded voice through headphones after a short delay induces stutter-like repetitions, prolongations, and blocks in speech. Lee explicitly referred to Norbert Wiener's book CYBERNETICS, published in 1948, as supplying the "fundamental concepts" to explain the effect. Black had called on the feedback principle to account for the reduction of voice level brought about by sidetone[1] amplification, and the raising of voice level which occurs in an acoustically dead environment.

However, as we have seen earlier, a feedback principle can be shown to be implicit in accounts of movement production from the time of La Mettrie, and, just as the physiological discoveries of Bell, Magendie, Claude Bernard, and Sherrington & Mott conferred empirical justification on this assumption, so Wiener's book, which collated existing theories of control and feedback arising out of computer science, provided mathematical justification and led to the explicit use of cybernetic formulations in many fields of study (Rose, 1971, p 17).[2]

It seems, then, that the only difference in explanations of speech production brought about by the 'cybernetic revolution' was a terminological one. In his first publication on DAF (1950a), Lee called it "side-tone delay", only later changing it to "delayed speech feedback" (1950b).

From 1950 onwards, it became de rigueur to refer explicitly to a feedback mechanism when writing about speech production. Black (1951) wrote that feedback monitoring has a <u>self-evident</u> role in speech acquisition:

> "The importance of the feedback monitoring system in speaking has not been fully determined. It is self-evident that it is indispensable in learning normal speech." (p 56)

---

1. 'Sidetone' is the pre-1950 term for what is now called 'auditory feedback'.

2. Sluckin (1954) puts back the collation of ideas about feedback and steady-state maintenance to an earlier date, 1940, and a different author, Ashby.

Black also expressed, in 1954, the optimism aroused by the possibilities of applying cybernetics to speech therapy:

> "When the story of sidetone is written we may lament that we did not declare a moratorium on much speech correction, voice and diction courses, etc. with CYBERNETICS until we found the nature of feedback in speech and how to cope with it." (p 143)

In the same year Fairbanks published his widely referenced article on the speech production mechanism as a servo system. It was Fairbanks who first proposed the use of the full term 'auditory feedback', as though it were as purely descriptive as the older, neutral term 'sidetone'.

There followed a spate of theoretical papers describing the speech production system in cybernetic terms (Peterson, 1954; Fry, 1954; Black, 1954; Peterson, 1955). Although this was apparently a new theoretical insight, the only research stimulated by it was further work on DAF in speech (Flanagan, 1951; Atkinson, 1952; Deutsch and Clarkson, 1959) and in other motor activities (Kalmus, Denes and Fry, 1955; Chase et al., 1959). Cherry and colleagues (1955, 1956) suggested that stuttering could be due to some type of perceptual delay analogous to DAF. Finally, some work was done on the filtering of sidetone (Atkinson, 1953; Malloy, 1953; Peters, 1955) which found that filtered sidetone <u>improved</u> the speaker's intelligibility.

It is interesting to note that the theoretical papers discuss exclusively the effects of DAF or of congenital deafness on speech, and never refer to Lombard's work or to the effects of adventitious deafness, although it is only the effects of <u>adventitious</u> deafness on speech which provide any support for the notion of auditory control.*

---

* But see Chapter Five

In considering the modern expressions of the notion that speech is controlled by self-hearing, we have been confined to what could be called in this context the non-medical tradition, simply because this body of writings does form a coherent whole in so far as authors writing within it tend to refer only to work by other non-medical writers. These writers tend to be unaware of the contributions which have been made by medical and physiological workers such as Bell, Magendie, and Claude Bernard to the general notion of sensory-motor control*, although Cannon's work on homeostasis is sometimes referred to. We have noted already that even Lombard's work tends to be ignored by the modern non-medical writers, although his work has always been known in the field of otology, since it stimulated the development of tests for deafness and auditory malingering.

(ii) Views Against

(a) The need to attenuate self-hearing

There is also a medical tradition of work on the role of self-hearing in speech, which naturally enough has concentrated on the physiological links between speech and hearing. Writers in this tradition agree with the non-medical writers that there is an interaction between speech and self-hearing, but their view of the interaction is radically different; it is that self-hearing must be attenuated during speech.

Von Bekesy (1949), following Barany (1938) points out that a vocaliser (human or animal) capable of producing intense sounds, may need to be protected from the vibrations of its own sound-making, and he derives the anatomical structure of the middle ear from this supposed function. He illustrates this point in the frog and the rooster: both have the same mechanism for this

* Chase, who has worked on DAF, is an exception - but he is medically qualified.

function, in that the opening of the jaw automatically closes the auditory canal.

In addition to the anatomical evidence, there are numerous reports of vocalisation being accompanied by middle-ear muscle activity which alters the sound-transmission properties of the middle ear, both in animals and in humans (Carmel and Starr, 1963; Shearer and Simmons, 1965; Karlovich and Luterman, 1969). Jen and Suga (1976) point out that, in echolocating bats, the middle ear muscles are activated about 3 milliseconds after the laryngeal muscles. They suggest that this "attenuates vocal self-stimulation and improves the performance of the echolocation system." Bosatra, Russolo and Semeraro (1975) elicited contraction of the tympanic muscle by electrical stimulation of the tongue in humans. Gaynor (1974) reports that loud humming can induce substantial pure-tone threshold elevation, the maximum sensitivity shift being observed in the region of maximum energy of the speech signal.

Borg and Zakrisson (1975) give a different interpretation of this middle-ear activity during vocalisation, suggesting that it functions to prevent the possible masking of other sounds which may be brought about by one's own sound-making. The two interpretations are not, however, in conflict.

The fact that middle-ear contraction in humans often occurs immediately before speech production, and hence is not merely an acoustic response, is taken by Borg & Zakrisson (1975) to indicate that the stapedius muscle, at least, "can be innervated from the central nervous system as a part of the vocalisation process". (p 325)

McCall and Rabuzzi (1973) also hypothesise, on the same grounds, that "the tympanic muscles are activated relatively simultaneously with the speech musculature by neural impulses directly from the motor cortex" (p 56).

Shearer (1966) states that this middle-ear reflex is elicited only in relation to speech movements of the articulators, and not, for instance, by random tongue and jaw movements, leading him to conclude that "the production of speech, rather than the motor movement itself, is the key factor in the reflex" (p 1280).

It has been found by other workers that contractions of either the stapedius or the tensor typmani muscles may accompany non-speech motor activity such as lifting the head, yawning, swallowing, blinking, and gross body movement (Salomon & Starr, 1963), but Borg & Zakrisson suggest that such studies have not used sufficient precautions against electrical interference from activity in neighbouring jaw and face muscles.

A more direct piece of evidence for the role of the middle ear in attenuating the loudness of one's own voice is provided by Melnick (1965), who did an autophonic matching experiment with patients who had undergone stapedectomy (ie asking them to match their own voice loudness against the loudness of external sounds). He found that the apparent loudness of their own voice to themselves had increased as a result of the operation.

Shearer (1966) found a rough correlation between middle-ear muscle contraction and voice intensity in normal speakers, but that the correlation was much less close in the case of stutterers, suggesting that stutterers may be less well defended from their own voice intensity than normals. McCall (1973) has found middle-ear dysfunction of a similar kind in patients with spasmodic dysphonia, although he attributes both the voice and the middle-ear disorders to impairment of the extrapyramidal motor system, rather than to any direct effect of middle-ear functioning on voice.

Cherry and Sayers (1956) reported that stuttering could be suppressed if the very low frequencies of the stutterer's sidetone were masked, and this finding taken in conjunction with Shearer's suggests that stutterers may experience less effective attenuation of the low frequencies of their sidetone than normal speakers.

There are other indications that the attenuation of low frequencies can facilitate some auditory tasks. It is well known that the fundamental frequency (F_o) of the human voice need not be physically present in order to be perceived; it is inferred from the harmonics. Klatt (1973) has demonstrated that the discrimination of jnd's of the F_o of synthetic vowels was actually improved if the F_o itself was removed by high-pass filtering, and he concluded from this that "the fundamental component is not involved in the detection of changes in F_o" (p 8).

Finally, this apparent 'dispensability' of low frequencies and the relation of vocalisation movements to the transmission properties of the middle ear may be linked to the "ancient belief of sailors" that opening the mouth wide enables one to hear better (Tournier, 1974, p 70) - could this practice, reminiscent of von Bekesy's example of the frog and the rooster, attenuate the low, masking rumble of a sailing ship? A possibly related observation is that people tend to open the mouth when straining to hear, and when surprised or puzzled, though Coiter, in the sixteenth century, suggested that a deaf person may hear better with the mouth open because air conduction along the Eustachian tube can then occur (Stevenson and Guthrie, 1949).

(b) The Cochleo-Phonatory Reflex

Another branch of the medical tradition, mainly a continental one, has gone further than the medical writers previously discussed (who have emphasised the need to attenuate one's own voice) and have insisted that auditory effects on voice are not functional at all, but merely represent incidental nervous stimulation of the larynx from the cochlea (Husson, 1962; Molinari and Pivotti, 1962). Thus the Lombard effect is described by Garde (1965) as "non un effet d'assourdissement, mais bien un effet de stimulation"* (p 632).

The evidence for this claim dates from 1951, according to Husson (1962), when the following observations were carried out by Husson and others. If a "pianissimo"

---

* "not an effect of deafening, but rather an effect of stimulation."

tone was fed into the ear of a subject simultaneously vocalising on the same note "piano", the vocal chord ipsilateral to the stimulated ear was seen to stiffen. If the intensity of stimulation was increased, the other vocal chord was also affected. If the frequency of the introduced sound was slightly different to that of the sound being phonated, the regularity of vocal-chord vibration was impaired. This effect was named the cochleo-recurrential reflex (the recurrent nerve is a branch of the vagus, the tenth cranial nerve, which innervates all but one of the intrinsic laryngeal muscles); Garde (1965) calls it the cochleo-phonatory reflex. According to Husson (1962) the effect has since been observed many times.

Husson and his co-workers also examined the vocal chords of deaf subjects who possessed speech, and found that they were in a condition of permanent hypotonia. Furthermore, they found that a patient with unilateral cochlear deafness had a hypotonic vocal chord on the same side as the deafness. However, no attempt is made to derive the quality of 'deaf voice' from this state of the vocal chords.

In order to support his explanation of the Lombard effect as one of stimulation, rather than deafness, Garde (1965) carried out two informal experiments. In the first, he simulated deafness in a normal subject with the use of ear-plugs, and found only a diminution of voice level. However, he did not carry out any objective measurement of voice level, and his result conflicts with those of Rubenovitch and Pastier (1938), Kryter (1946) and with my result in Experiment II of an inaudible, but measurable and statistically significant <u>increase</u> of voice level with ear plugs.

In the second experiment, Garde applied a low-frequency vibration to the speaker's mastoids, a procedure which produced bone-conducted auditory stimulation while at the same time allowing the subject to hear both his own voice and the whispered voice of the experimenter. This produced"des modifications de la voix"(changes in the voice) which are not further specified by Garde.

However, although Husson's group and Garde have repudiated Lombard's conclusions, they have not explicitly addressed themselves to the general question of the role of self-hearing in speech, nor have they suggested any function for the cochleo-phonatory reflex.*

There seems no reason to deny the findings of Husson's group, although Garde's two experiments claiming to show that the Lombard effect acts as a stimulation effect are inadequate, at least as reported. The absence of any explanation of their findings leads to the conclusion that they have not effectively undermined the possibility that self-hearing has a role in speech production.

(c) The Intelligibility-Conserving Theory

A much stronger attack has come from Lane and Tranel (1971). They have reviewed a large body of literature on the so-called auditory feedback effects and the role of self-hearing in speech, and have argued persuasively that these effects can be re-interpreted simply as efforts by the speaker to remain intelligible, rather than as deficits in performance. In other words, speakers do not listen to *themselves*, but to the acoustic environment in which they are attempting to communicate, and they adjust their speech accordingly. This argument is supported impressively by the facts that (i) the Lombard effect is stronger when the speaker is actively trying to communicate, rather than merely reading out word-lists; (ii) none of the previously-known effects, apart from the important exception of DAF, actually result in impaired intelligibility, as the deficit explanation should entail.

Lane and Tranel do not consider DAF to fall outside their explanation; on the other hand they offer no explanation of why DAF speech is so distorted (its poor intelligibility is so evident that it requires no experimental confirmation).

---

*Jen & Suga (1976) found larynx muscle activity after sound stimulation in bats, and suggest that this acts as negative feedback to stabilise the accelerating emission of pips in the last moments of echolocation.

Lane and Tranel's theory is my reference-point throughout the rest of this thesis; it will be picked up in detail as I give closer consideration to the evidence for and against self-hearing in speech production in the next chapter, and the experiments to be reported are directed at specific aspects of their argument.

CHAPTER TWO

DETAILED EXAMINATION OF THE EVIDENCE

We have seen in the preceding chapter that the modern views in favour of the role of self-hearing in speech have, at least since 1950, been largely theoretical, in the sense that they have used cybernetic theory as an illustration of something which was simply assumed to be true - namely that speech is controlled by auditory feedback. Clearly, these writers were not defending the role of self-hearing. However, especially since the publication of Lane and Tranel's paper, the notion that self-hearing has a role in speech production needs defence, and consequently closer attention must be given to any sources of evidence which may support or refute it.

In the first chapter, we looked briefly at the sources of belief in the role of self-hearing which existed before the Lombard effect was reported in 1909. In this chapter, we will consider some of these sources of evidence, which are still occasionally appealed to, in greater detail, and the next chapter will reconsider Lane & Tranel's theory in the light of conclusions from the present chapter, and of our own experimental findings.

1. The Evidence from Deafness

The failure to acquire normal speech as a result of congenital deafness

Little need be said on this point. Although many authors suggest that the speech difficulties of the congenitally deaf are due to their inability to hear their own efforts at speech (Davis, 1951; Fry, 1954; Lenneberg, 1964; Luchsinger and Arnold, 1965; Harris, 1970) it is clear that the inaccessibility to a deaf child of normal speech models is sufficient to explain the failure to acquire normal speech. The fact that, even with modern visual aids to represent speech patterns to the congenitally deaf, their difficulty persists, does not argue against this point; for we do not yet know whether these visual representations contain all the information in the corresponding acoustic waveforms.*

The deterioration of speech following adventitious deafness

At first sight, the common observation that speech deteriorates after adventitious deafness ('AD') seems to offer the best possible evidence for the role of self-hearing in the regulation of speech.

However, there are apparently no systematic longitudinal studies of how speech deteriorates in this condition, and, somewhat surprisingly, there is disagreement among authorities as to whether or not deterioration actually occurs:-

> "The usual changes long-recognised as typical of the speech of a totally deaf person are a rise in pitch and intensity of voice, and a flattening of intonation." (Stanton, 1958, p 38)

> "The onset of deafness after speech has developed does not usually interfere with the ability to speak, except that some will tend to shout owing to difficulty gauging the loudness of their voice." (Espir and Rose, 1970, p 113)

This disagreement may be due to the possibility that some individuals are more affected than others, or that different writers have different criteria for what constitutes "interference with the ability to speak", or to the fact that speech deterioration is not immediate after the onset of deafness, but occurs gradually, at least in adults (Dalby, 1873; Carhart, 1947; Harris, 1970).

*It is interesting to note, however, that while laughter in the congenitally deaf is acoustically normal (Lenneberg, 1964) their breathing is noisy; and some deaf children fail to acquire the normal chest register after puberty (some boys keeping a falsetto voice) (Luchsinger & Arnold, 1965). This indicates that breathing and voice quality may have learned components, while laughter need not.

Luchsinger and Arnold (1965) give the following account of AD speech. Respiration may be uneven, intensity may be wrongly used (both in regard to general speech level and acoustic stress) and intonation may be incorrect or lacking. Articulation may be characterised by general slurring - "the fricatives are especially vulnerable because they are motorically difficult and require precise auditory monitoring" (p 635). There may be failure to differentiate voiced and unvoiced consonants, and the vowels tend to collapse to a mid-point.

Dalby (1873) gives a similar account of 'indistinct' and 'thick' speech in AD children, but such observations do not lead him to the usual conclusion that self-hearing is necessary for speech, but rather to an important alternative explanation of speech impairment following AD - that the important loss is the sound of others' speech, not the sound of one's own:-

> "the child not hearing what is said around very soon forgets the knowledge that it has not practised itself in exercising for a long enough period to make speech a confirmed habit. The same influences which make a child in this manner dumb will cause a child to lose one language while it is acquiring another. Thus a boy or girl of four or five years of age who has been brought up in India with a native nurse and taught as a first language Hindostanee will in six months if it is brought to England and does not hear the language spoken have completely forgotten it." (p 201)

In other words, Dalby suggests that, for normal speech to be maintained, there must be continued speech input, or access to auditory models of speech.

The alternative explanation hinted at by Dalby will be considered in detail later (Chapter Five); for the moment I only want to make the point that it is, prima facie, a plausible alternative explanation.

Some findings by Penn (1955) on the other hand are difficult to interpret as other than indicating a function of self-hearing in normal speech. He compared voice and speech patterns in groups of perceptively and conductively AD American veterans, and found distinct voice and speech patterns associated with each type of deafness. (In conductive deafness, some hearing through bone-conduction is preserved.) There were more deviations overall in the perceptive group. "Audible breathing, unconscious phonation, omission of high-frequency consonants in consonant clusters, general vowel confusion, and distortions of the phonemes "f, v, k, and g" were frequent in the perceptive group, but absent in the conductive group. The perceptive group had a greater incidence of "excessive volume, nasal quality, strident quality, monotonous pitch, poor mobility of the articulators, and distortions of the phonemes r, θ, ð, s, l, č, ǰ, ž, and š" and the conductive group had a greater incidence of "denasal quality, weak volume, unvoiced,weak or omitted final consonants, and distortions of the phonemes n and m".

Lane and Tranel's only comment on speech deterioration after AD is that speech cannot be maintained indefinitely without the "public loop", ie knowledge-of-results type of feedback. They of course deny that the 'private loop' (sidetone) is in any way involved.

In concluding this section, it would appear that Penn's findings are the only solid evidence to indicate that speech deterioration after AD occurs because of the loss or distortion of sidetone. It should be pointed out however, that even these findings are open to another interpretation. It is possible that AD speakers modify their speech in accordance with the (apparently changed) speech they hear - in other words, that their speech is a faithful reflection of the speech (of others) that they actually hear. Whether such a feat could be achieved without the use of auditory self-monitoring is, of course, the question at issue, although it would seem that the fact of not being able to hear oneself accurately in these circumstances would diminish the possibility of self-monitoring being involved.

2. The General Model of Sensory Inflow Control of Movement, and its Relevance to Speech

Although the notion that voluntary movement is controlled by sensory feedback has found widespread acceptance*, it is far from being rigorously established, and is coming more and more to be seriously questioned.

Current theories of skilled movement (see Glencross, 1971, Connolly, 1977) while acknowledging the demonstrations of impaired performance where sensory inflow is absent, delayed or distorted (predominantly, in fact, these are demonstrations of DAF in speech) also point out the difficulties in accepting that skilled movement is controlled by sensory inflow.

A major difficulty, early pointed out by Woodworth (1899), and by Craik (1947) and Lashley (1951), is that, particularly for highly skilled sequences of movement, there does not appear to be time for sensory feedback from each unit of movement to be processed before the next unit must begin. Furthermore, the extreme complexity of a skill such as speech, involving very many muscles simultaneously and "several hundred events every second" according to Lenneberg (1967, p 92) while not ruling out the possibility of control by sensory inflow from all these events, does at least make the possibility seem uneconomic and implausible. Lenneberg concluded that "There must be some automatisms - whole trains of events that are 'preprogrammed' and run off automatically". (p 92)

---

* "If one considers the results of a large number of neurophysiological studies of sensorimotor function, the overwhelming impression is that closed-loop control is a universal property of behaviour." (MacNeilage and Ladefoged, 1976, p 106)

Speed and complexity, then seem to argue against the use of sensory inflow to control movement. Such arguments have led to attempts to demonstrate the dispensability of sensory inflow - most notably the demonstrations of Taub and Berman (1968). They have reported work which demonstrates "the startling results ... that finely co-ordinated movements and learning ARE possible in the absence of somatic sensory feedback" (Freeman, same volume). This work appears to have superseded the early classic experiment by Sherrington and Mott (1895) which demonstrated that, in monkeys, deafferentation of a limb resulted in loss of purposive movement in the limb, and from which it was concluded that somatic sensation is necessary for voluntary movement. Taub and Berman reasoned that if purposive movement could be demonstrated following "the interruption of all relevant spinal reflex arcs", then the classic conlusion would be undermined.

Accordingly, they deafferented a single forelimb in two groups of monkeys, one of which had previously been conditioned to flex the limb to avoid shock, while the other group were conditioned post-operatively. Although there was an initial deficit in the response of the pre-conditioned group after the operation, they were reconditioned back to acquisition criterion, and all the pre-operatively naive animals were able to learn the response.

To exclude the possibility that other sources of feedback were being substituted, the next strategy was to remove these. The response-terminated buzzer of the first experiment was replaced by a brief click, and the gross limb flexion movement (which could give rise to skin distension beyond the deafferented limb, or to stimulation of the middle ear organ of balance) was replaced by a grasp response. Although the animals were permitted to view the experimental arrangement before each training session, during training itself the animals could not see their limbs. Every animal learned the grasp response, and was able to grasp as strongly as normal animals.

A final experiment deafferented both forelimbs, and the conditioning results were as before. It had always been observed previously that, in the free situation, animals with a single deafferented limb failed to use it purposefully unless the intact limb were restrained, but with both forelimbs deafferented, there was a gradual restitution of normal functioning, taking from 2 to 6 months. Taub and Berman claim that, after maximal recovery, "the degree to which the movements of these animals ... approximated normal patterns of movement was truly striking and cannot be overemphasised". Moreover, "these animals displayed as large and well-co-ordinated a range of movements with a blindfold on as with vision unobscured". (p 177)

These studies must throw new light on the role of sensory inflow in movement, and in the words of Taub and Berman themselves, "That such research had not begun earlier is really quite surprising in view of the extensive theoretical use that has been made of proprioception by psychologists ... Instead we have had, almost uniformly, the tendency simply to assert the significance of proprioception, as if its omnipresence assured its relevance." (p 174)

Of course these experiments were mainly concerned with intrinsic inflow, whereas audition provides extrinsic inflow from speech. Even so, by their control of visual inputs from deafferented limbs, Taub and Berman have also seriously brought into question the necessity of extrinsic inflow in the control of movement.

The issue is not quite as simple as this, however. The effects on performance of experimentally manipulating sensory inflow still need to be explained (and these effects include the enhancement, as well as impair ment of performance, in some cases). One point relevant to such an explanation arises out of Taub & Berman's own work.

Glencross (1977) has emphasised that while there is striking motor recovery in the de-afferentation studies, nevertheless "the subjects are never 'normal' and the elegance of movement is lost" (p 24). Glencross goes on to define 'elegance' in terms of co-ordination of successive units, fine control and precision, precisely those attributes which Connolly (1977) uses to characterise a skill. Abbs & Eilenberg (1976) also review experiments in which intrinsic afferent blocking resulted in loss of fine motor control. De-afferented monkeys, then, can still perform motor actions - but there is a case for saying that they do not perform them skilfully. Just as adventitiously deaf speakers may still speak, but they do not speak elegantly.[1]

It is clearly possible, then, that sensory inflow is used mainly for sequencing, fine control, and precision of movement. Sequencing in speech is partly a matter of co-articulation, and it has been shown that co-articulation features are absent in DAF speech, and in stutters (Rawnsley & Harris, 1954). It can also be easily demonstrated that sensory inflow is needed in order to provide absolute, as opposed to relative, anchor-points for performance. For instance, if one attempts to write with closed eyes, the individual letters can be formed normally, but the line of writing may not be straight across the page - it may stray upwards or downwards. Similarly, if one attempts to draw a 4-pointed star with closed eyes, it is very much a hit-and-miss matter whether one manages to close the figure:  Without visual inflow, one cannot keep track of the starting-point.[2]

1. Glencross also raises the possibility that movement control mechanisms may differ in monkeys and humans - indeed the evidence indicates that central control is more predominant as the evolutionary scale is descended, as one would expect given that humans are pre-eminently the 'learning' species.

2. The fact that one can write using different instruments and on surfaces of different slope has led Arbib (1972) to suggest that action programmes stored in the brain specify relative, but not absolute motion orientations.

This point will be particularly relevant when we come to discuss the role of self-hearing in the relative vs. absolute control of voice level (Experiment VI).

If we now confine our attention to extrinsic inflow - ie, for example vision in the case of Taub and Berman's monkeys, vision in the case of handwriting, and hearing in the case of speech, then the point just made could also be expressed in terms relevant to speech by suggesting that the suprasegmental features (intonation, intensity in speech; alignment in handwriting) may be under extrinsic inflow control, while the segmental features (articulation in speech; individual letters in handwriting) are not, or at least not to the same extent. (This is not to suggest that articulation in speech may not be under intrinsic inflow control - ie tactile or kinaesthetic. Indeed Lane and Tranel, while dismissing any role for auditory inflow in speech production, nevertheless claim that speech loudness is regulated entirely by 'vocal effort' - ie by intrinsic sensory inflow.)

It is just these suprasegmental features of speech - intonation, intensity - which vary continuously in a gradated, rather than segmental way, which would be expected to make use of absolute anchor points in performance. At the same time, the suprasegmental features are long enough in duration for auditory inflow to be usefully processed, whereas individual speech sounds are not:

> "(A) theoretically important matter that has perplexed the present writer is the nearly instantaneous effect of side-tone feedback on the relative levels of different vowels produced with 'the same' vocal effort ... It is not unusual to discover that a speaker will alter his vocal level on the basis of information derived from side-tone feedback, but it is hard to imagine a perceptual process that can act so fast that the peak SPL of the medial vowel of a stressed monosyllable is modified
>
> .. .. ..

.. ..
by information fed back from initial segments of the same vowel. The solution of this puzzle is left as an exercise for the reader." (Allen, 1971, p 1840)

"It seems unlikely that moment to moment auditory feedback plays an important role in control of running speech, largely because speech movements for the most part precede their main acoustic effects in time, and the firing of the motoneurons controlling the muscles largely precedes the movement." (MacNeilage, 1972, pp 43-4)

"Speakers may be able to correct some of the longer speech sounds, vowels and some fricative consonants, as they are articulating them. Other sounds of short duration like the stop consonants are completed before speakers can perceive them and can alter their articulation. Because utterances are usually as long as several words in length, speakers can monitor the loudness, pitch, stress, and rate of articulation and can alter them as well as the overall precision of articulation." (Daniloff, 1973, p 183)

It is interesting to note that Kozhevnikov and Chistovich (1966) have adopted for rate in speech (a suprasegmental feature) the same solution that Arbib (1972 - see footnote 2, p 30, this thesis) has adopted for varying surface slope and instrument in handwriting. All of the suprasegmental features are continuously-varying rather than discrete, and require specification, at some point in performance, in absolute as well as relative terms - ie some specification of anchor points as well as specification of the relative values of units in unit-to-unit sequencing:

"If the rate of speech can change continuously over a sufficiently wide range (can employ any value within these limits) it is unrealistic to assume that for each possible rate an individual articulatory program is constructed. This simply would be difficult to do and, in addition, it would be clearly

.. .. ..

> .. ..
> uneconomical. It is considerably more natural to make another assumption, namely that the rate in no way figures in the articulatory program." (Kozhevnikov and Chistovich, 1966, p 77)

It is possible then that the correct attainment of such anchor-points, or absolute values, in speech is dependent on the processing of extrinsic sensory inflow - ie on self-hearing, in the case of at leastintonation and intensity (the control of rate seems less straightforward). This is not to imply that no components of the suprasegmental features are under central control - rather that, as Kozhevnikov and Chistovich suggest, the sequencing of relative values is controlled centrally, while the absolute values are arrived at through comparison with extrinsic sensory inflow.

While we have underlined the usefulness of absolute anchor-points for suprasegmentals, it was not meant to imply that such anchor-points are never used in articulation. Clearly, in spite of a degree of permissible variation in articulation (dialect, idiolect) some external check is needed even here. This possibility that articulation is less under extrinsic control than the suprasegmentals is supported by the fact that, after AD, the suprasegmentals deteriorate more rapidly than the segmental features of speech. Further, the possibility of different control systems for suprasegmentals and segmentals is lent added credence by the fact that different muscle systems are involved in their production. Voice level and intonation are produced by the respiratory and laryngeal musculature, while articulation is carried out by the orofacial musculature (Harris, 1970).

It is interesting to note also that Lee himself, who first demonstrated the DAF effect, suggested that 'voice' is under auditory control,while articulation is under kinaesthetic control,and he suggested this because he thought, apparently, that kinaesthetic inflow could be processed more rapidly than auditory inflow:

> "The elements of articulation are commonly accomplished at a rate of 14 per second while distinct utterances of the vocal chords alone can be performed at only a quarter of this speed." (Lee, 1950b)

We have just discussed, then one way in which speech could be under extrinsic sensory inflow control - namely, that the suprasegmental features, at least intonation and intensity, are monitored extrinsically and compared with some internal model of an absolute value of frequency and intensity.

Further consideration of the ways in which sensory inflow could be used in the control of movement involve the notions of loading and intermittent control.

MacFarland (1971) has pointed out that closed-loop control is found in cases where disturbance of output may be expected to occur - for instance in control of the limbs, which are subject to loading and disturbance. Eye movements, on the other hand, do not normally have to contend with loading or disturbance, and apart from some types of visual tracking, eye movement control is not, as far as is known, carried out using sensory inflow.

Bernstein (1967) has made essentially the same point in underlining that there are many instances of locomotion in which the independent forces involved are unforseeable (running over uneven ground, jumping onto an elevation, swimming through waves) - "and because of this they cannot be overcome by any sort of stereotyped movements directed solely from within". (p 115)

On this consideration, we would not expect speech to be under closed-loop control, for it is, like eye-movements, one of the few instances of movement which are not normally subject to loading or the overcoming of independent forces. The performance conditions for speech change rarely - only when one is simultaneously speaking and eating, smoking or holding a thermometer in the mouth, or after being anaesthetised by the dentist. In other words, when one is simultaneously speaking and holding some object in the mouth, or when some of the articulators are not in their normal condition.

Suppose, then, that speech is <u>normally</u> a predominantly open-loop skill, but comes under closed-loop control when there is loading or unpredictability. (Lane and Tranel have indeed suggested that experiments which interfere with speech sidetone actually <u>induce</u> closed-loop control by making the performance conditions difficult, rather than demonstrate that this is the normal mode of speech control.) Even in the case where control is open-loop, there must, as Poulton (1957) points out, be occasional or <u>intermittent</u> checking against the environment:

> "This is because with the passage of time the performance tends gradually to drift in the positioning and/or timing from the optimal requirements. The less exacting the requirements are, and the more practised the subject is, the longer will be the time for which the skill can proceed without a check." (p 474)

Craik (1947) had also made this point, and provided empirical evidence, in the form of pen-recordings of manual tracking, for intermittent corrective movements:

> "The human operator behaves basically as an intermittent correction servo ... The intermittent corrections consist of 'ballistic' movements ... In playing musical instruments, typewriting, sending morse, etc., complicated patterns of movement are executed at a rate which would be impossible if they were continuously governed by the value of the misalinement (sic), with the inevitable reaction-time lag. Apparently, they must be individually performed, triggered off ballistically, and the sensory feedback must take the form of a delayed modification of the amplitude of subsequent movements. Sensory control, in other words, alters the 'internal gear ratio' or amplification of the operator with a time lag and determines whether subsequent corrective movements will be made; it does not govern the amplitude of each individual movement while it is being made." (p 56)

This 'intermittent' solution is in fact the one usually adopted by writers in speech control who accept the time and complexity objections to continuous monitoring of sensory inflow (Kezhevnikov and Chistovich, 1965; Daniloff, 1973).

It should be pointed out here that extrinsic sensory input, whether continuous or intermittent, gives us just as much information about the external environment as it does about the success of performance. Similarly, it is surely also true that in many experiments which vary sensory inflow, and claim to be varying 'feedback', the whole environment (at least in that sensory modality) is being varied, not simply the sensory inflow from movements, eg Stratton's (1896, 1897) experiments with inverted lenses, Held's (1965) experiments with prisms, and indeed speech experiments with varied sidetone. This point is analogous to Lane and Tranel's in that they claim it is the acoustic environment which controls speech, and not self-hearing.

Summing up this section, I have argued that neither the time and complexity objections to the sensory control model, nor Taub and Berman's demonstrations of relatively normal movement in de-afferented monkeys necessarily lead to the conclusion that sensory control does not occur. Rather, I have reviewed evidence - the inelegance and absence of fine control in de-afferented monkeys, and the indispensability of extrinsic sensory input in skills such as writing and drawing - which indicates that sensory input probably is needed in sequencing and setting absolute values and anchor-points for performance. Such absolute values and anchor-points would appear to be necessary for the correct production of suprasegmentals in speech, and furthermore the duration of suprasegmentals is great enough to nullify the time objection to continuous monitoring. However, it was pointed out that speech as a skill rarely has to contend with loading or unpredictable performance conditions, and so continuous closed-loop control would appear to be unnecessary. Nevertheless, there remains the probable necessity for intermittent sensory monitoring in order to re-set performance values against environmental values (although it was pointed out that any monitoring system is as much a means of checking on environmental values as of checking on performance values).

Clearly, then, the general model of the control of movement by sensory inflow is not one which applies universally to all aspects of movement. In the following sections, the evidence for the control of speech by self-hearing will be examined in more detail in relation to the effects of variation in side-tone and acoustic environment on speech.

3. Sidetone Effects on Voice Level

(i) The Lombard Effect

As I have already pointed out*, the Lombard effect was the first experimental demonstration of the effect on voice of varying a speaker's sidetone, and it has been accepted as clear evidence for the regulatory role of self-hearing in speech, up to (and beyond) Lane and Tranel's review. The effect has also been used, like the DAF effect, as a general example of the negative consequences for performance of removing feedback (Greenwald, 1970; Beloff, 1973). The effect could just as well be interpreted, however, as one of varying the acoustic environment. Furthermore, if the effect itself is considered carefully, it cannot be seen simply as a decrement in speech performance. There is some ambivalence on this question in Lombard's own accounts. On the one hand, he explains the effect as an effort by speakers to hear themselves above the noise - "de s'entendre mieux soi-même" ("to hear oneself better"; Lombard, 1911, p 101). This explanation, of course, presupposes that one needs to hear oneself during speech. On the other hand, the effect is said to be a loss of auditory control over voice level: "Ce

---

*p2

phénomène paraît résulter de la suppression brusque du contrôle auditif par le sujet lui-même sur l'intensité des sons émis pendant la phonation normale" ("This phenomenon appears to result from the subject's sudden loss of auditory control over the loudness of sounds emitted during normal phonation"; Lombard and Baldenweck, 1915, p 502).

Right from the beginning, then, there is ambivalence as to whether the effect demonstrates loss of control or not. Yet it is clear that some aspects of even voice level control are not lost during the effect, since what speakers do is to raise their voice level systematically with the noise level. Furthermore, if the effect were simply a loss of control over voice level, speakers would be just as likely to speak softer with noise, and this never happens.[1]

The only sense in which the effect can be called a decrement in performance is that the speech produced is simply too loud for the communicative or acoustic setting in which it takes place, unless (and this is an important possibility which will be picked up later[2]) we accept that speakers always assume that their listeners are subject to the same acoustic conditions as themselves. At present, however, we have no evidence for this possibility, and if we agree that Lombard speech is simply speech which is too loud, the only way in which this can be described as a performance decrement (bearing in mind the retained ability to systematically shift voice level) is to say that the performance anchor-point (or modulus, in psycho-physical terminology) has been unnecessarily shifted upwards. In other words, the appropriate absolute quantities cannot be attained in the absence of sensory inflow from performance, which matches

---

1. Except in some psychiatric patients (Rubenovitch and Pastier, 1938) and this is their only response.

2. (p 58)

performance with environment characteristics. This seems to be one possible account, yet even with this the systematic nature of the remaining performance is puzzling.

What, then, of the other explanations of the effect which have been offered? Apart from Garde's, already discussed*, the only alternative is that of Lane and Tranel, and they explain it as an attempt by the speaker to remain intelligible <u>for an audience</u>, rather than for oneself. In other words, one speaks louder so that others may hear, not so that one may better hear oneself, and the response is to the <u>presence</u> of noise, rather than to the <u>absence</u> of sidetone. This view can explain the systematic raising of voice level with noise, and is also supported by the fact that, when there is a listener being communicated with, the slope of the Lombard function (voice level plotted against noise level) is steeper than when the speaker is merely reading word-lists into a microphone.

There is other evidence to support this account. Louder speech is more intelligible speech (barring the upper limits of voice level; Pickett, 1956, 1958) and when running experiments on the Lombard effect one is impressed by subjects' improved ability to project their voice, particularly some women whose normal voice is timid and breathy. Mahl (1972) has described this striking change in subjects' self-projection under the Lombard effect as a kind of ego release, since it may include improved verbal fluency and extended free association. These are effects which have also been found by Holmes and Holzman (1966) and Klein and Wolitzky (1970).

Speech-masking noise has also been used as a therapeutic aid to improve speech (and one could call it speech intelligibility) in speech disorders as varied as aphasia (Birch and Lee, 1955, and Birch, 1956 finding both improved oral reading <u>and</u> word-finding), hysterical mutism (Lombard and Baldenweck, 1916-17), dysphonia (Luchsinger and Arnold), and stuttering (Cherry, Sayers and Marland, 1955; Shane, 1955). Furthermore, the

---

* Pp 20-21

stutter-like disfluences which occasionally occur in normal speech are also reduced by noise masking (Silverman and Goodban, 1972; Sherrard, 1975).

Garber, Siegel, Pick and Alcorn (1976) found that the magnitude of the Lombard effect was directly related to the speech-masking ability of the noise used. There is, then, considerable and varied evidence to suggest that the Lombard effect is really one of improved speech intelligibility.

Yet there are other aspects of Lombard speech which do not seem to be designed to enhance intelligibility. Firstly, the Lombard effect occurs in non-linguistic vocalization (sustained vowel) when there cannot be any intelligibility motivation (see Experiment III) though this could just mean that the response is purely automatic. Secondly, in some subjects Lombard speech is lacking in intonation, making it sound very much like deaf speech - an observation which Mahl, (1972) has also made. On the other hand, I have also observed that some subjects' intonation is exaggerated. Thirdly, Lombard speech in some subjects includes vocal fry phonation. Mahl (1972) interpreted such vocal fry events as proto-utterances which would normally be edited out by the speaker, but which partly escape under the cognitive confusion aroused by the noise, which sometimes prevents subjects from knowing whether they have spoken aloud or not. This interpretation does not, however, accord with enhanced intelligibility of the speaker. Fourthly, the speech may be slurred (Klein, 1965; Mahl, 1972). Finally, Mahl (1972) asked the sociolinguist William Labov to judge some of his speech samples, and he found that, under the Lombard effect, some speakers 'reverted to type' - ie reverted to a childhood or familiar accent and style of speech, often unrestrained and assertive. This finding again fails to support the notion of improved intelligibility, at least if we accept the sociolinguistic postulate that where communication takes place across any kind of social or psychological barrier, a closer approximation to the standard dialect and formal style will be used (Moscovici and Plon, 1966; Trudgill, 1974).

The possibility remains that the major characteristic of Lombard speech is improved intelligibility, with individual differences in respect of intonation, vocal fry, and reversion of speech style.

(ii) The Fletcher Effect

This is the effect whereby amplifying a speaker's sidetone causes a reduction in voice level. It was first reported in 1918 (Fletcher, Raff and Parmley), but no work seems to have been done on the effect between this date and 1949, when Lightfoot and Morrill replicated it. There are three important points to note about it. Firstly, it demonstrates of itself that one's own voice is a special auditory stimulus, since raising the level of any other auditory stimulus during speech would constitute 'noise', and induce a Lombard response. (Black, 1950B, found that loud tones of various frequencies would increase a speaker's voice level.) Secondly, it follows from the first point that one recognises the sidetone as one's own voice, and to that extent one must be monitoring one's own voice at least during the effect. Thirdly, it is difficult to interpret a reduction in voice level as an intelligibility-conserving effect, especially since Lightfoot and Morrill (1949), and Black (1950C) found that the reduced voice level was accompanied by reduced intelligibility, as Lane and Tranel do, unless, as for the Lombard effect, we accept that speakers assume that their listener is in the same acoustic environment as themselves. This is far from implausible, though to suggest that the assumption merely relates to the speaker's intelligibility somewhat demeans this particular sidetone effect. In reducing voice level, speakers surely would be concerned less with their own intelligibility than with their listener's comfort. It would appear, in other words, to embody a much less crude strategy than the Lombard effect does, on Lane and Tranel's interpretation.

A somewhat puzzling aspect of the Fletcher effect has been reported by Siegel and Pick and their colleagues (Siegel and Pick, 1974; Chang-Yit, Pick and Siegel, 1975; Garber, Siegel, Pick and Alcorn, 1976). They have found that the Fletcher effect is enhanced by mixing noise with the sidetone, that the enhancement increases with the noise level, and with the speech-masking ability of the noise (Garber et al, 1976), and furthermore that the noise level is a more powerful variable than the sidetone level itself in determining voice level (Siegel and Pick, 1974). The only interpretation offered by the similarly puzzled authors themselves is that the presence of the noise sensitises speakers to their sidetone* They also point out that previous experiments on sidetone carried out in the 1950's probably used amplifiers with high levels of internal noise, and so noise may always have been an unrecognised variable in the Fletcher effect. These authors did, however, still find a clear, if modest Fletcher effect when the amplified sidetone was <u>not</u> mixed with noise. In any case, the effect with noise is extremely difficult to interpret as an intelligibility-conserving one, since presumably remaining intelligible should take precedence over listeners' comfort when it comes to setting voice level in a noisy environment. The effect could, however, simply be an anomaly produced by giving the speaker conflicting cues about the acoustic environment.

(iii) <u>Sidetone Attenuation</u>

The converse of the Fletcher effect occurs when the level of sidetone is reduced. In this case, voice level increases, whether noise is mixed with the sidetone signal (Lane, Catania and Stevens, 1961) or not. Black (1950c) found that intelligibility increased as the level of sidetone was reduced.

---

* See p 157 this thesis.

It is not true, however, that sidetone level can be reduced indefinitely without adversely affecting communication. The first telephones did not incorporate a 'sidetone' circuit. This meant that people speaking into such telephones experienced effective attenuation in their sidetone, causing them to shout. For this reason, modern telephones incorporate a sidetone circuit but, interestingly, speakers are not normally aware of this except occasionally in relation to non-speech sounds such as breathing or oral clicks. Norbert Wiener writes that, presumably in oblivion of the early experiences,

> "dead-microphone transmission systems ... have actually been considered by the Telephone Companies, only to be rejected because of the frightful sense of frustration they cause, especially the frustration of not knowing how much of one's own voice gets onto the line. People using a system of this sort are always forced to yell at the top of their voices, to be sure that they have missed no opportunity to get the message accepted by the line." (1950, p 199)

Such telephones are analogous to the 'silent typewriter', an invention which failed because typists preferred to get auditory feedback, even at the cost of 'noise'.

Lane, Tranel and Sisson (1969) propose a combined interpretation of the Lombard, Fletcher, and Sidetone Attenuation effects which is based ultimately on the nature of the autophonic scale. An autophonic scale is set up by asking subjects to match the subjective loudness of an external sound with their own vocalisation, and it is found that the level of the vocalisation is just under half that of the external stimulus. Similarly, Lane, Tranel and Sisson found that in the Lombard, Fletcher and Sidetone Attenuation effects, the level of voice changes linearly with a slope of approximately 0.5, in relation to the level of sidetone or noise. The fact that, in all these cases, the 'compensation' by voice level for sidetone or noise level was only half-way is explained by them as follows: "the speaker matches changes in signal or in noise to keep the signal-to-noise ratio nearly constant, but he is misled by the disparity in the sensory operating characteristics of speaking and listening." (p 618).

(iv) Ear-plugs

There are two different effects of ear-plugs on the speaker's voice level, according to whether or not noise is also present. Some studies have been concerned with the practical question of how protective ear-plugs affect speech communication in noisy conditions. One such study by Kryter (1946) incidentally found that speakers who wore ear-plugs in noise spoke less loudly than speakers in the same conditions with unprotected ears. A similar study by Lower and Martin (1976) found the same result, and also that ear-plugged speakers were less intelligible than unprotected speakers in noise. (In this study, the listeners who found ear-plugged speakers less intelligible were in the same background noise themselves, but another panel of listeners who were not in the same background noise found no intelligibility difference between ear-plugged and unprotected speakers.)

When ear-plugged speakers are not in a noisy environment, however, they have been found to speak louder than when not wearing ear-plugs (Rubenovitch and Pastier, 1938), and also more intelligibly (Black and Tolhurst, 1956).

The findings in noise are straightforwardly explained as follows: in noise, the ear-plugs act to reduce the noise level for the speaker, hence what occurs is effectively a weaker Lombard effect (weaker than for speakers in the same noise, but without ear-plugs). The findings in quiet, however, conflict with what might be expected,since when the ears are occluded the bone-conducted sidetone seems louder than usual. Since, at the same time, air-conducted sidetone is attenuated, it is possible that if sidetone is being used at all, it is the attenuation of air-conducted sidetone which is being responded to when ear-plugged speakers raise their voice level in quiet. A further factor may be that, as von Bekesy (1962) has pointed out, it is easily shown that 'internal noise' is increased by placing fingers in the ears.

Of two studies which have found reduced vocal level with ear-plugs worn in quiet, Hebb, Heath and Stuart (1954) kept their subjects ear-plugged over a period of two weeks, possibly bringing other mechanisms into play; and Garde's (1964) study, finding either no effect at all or reduced voice level, was based on his own subjective assessments of voice level. Given that the order of increase of voice level with ear-plugs in quiet is only 3-4bD, subjective ratings would not be reliable.

The complexity of the findings relating to ear-plugs reveal that what is really at issue here is whether ear-plugged speakers set voice level according to what is heard of the environment, or according to what is heard of sidetone (or, if both are involved, the relation between them). Lower and Martin (1976) claim that, when ear-plugs are worn in noise, the apparent noise is reduced, but they speculate that one's own voice loudness is hardly affected (presumably compared to speaking in noise without ear-plugs), because of the compensating increase in loudness of the bone-conducted sidetone (compensating, that is, for the reduction in air-conducted sidetone). This would suggest that, in this case, it is the environment which matters. However, their own finding that ear-muffs gave more attenuation than ear-plugs, yet had no different effect on voice level ("... just as much, probably because they enhance bone conduction of one's own voice") suggests the reverse - that what matters in setting voice level is not the environment, but bone-conducted sidetone.

In other words, an ear-plugged speaker in noise receives about the same level of sidetone as a speaker without ear-plugs, but hears less of the noise, and so produces a weaker Lombard response. Then, the factor bringing about the change must be the environment. On the other hand, we know that the attenuation (environment information) given by ear-plugs and ear-muffs differs, and that in both conditions voice level is the same. In this case, we can infer that the levels of bone-conducted sidetone must differ.

It may help to sort out the relevant factors and their likely values and effects if we cast them into a Table (Table 1). For comparison, the bottom half of the table includes the Fletcher effect and Sidetone Attenuation effect.

Giving the arbitrary value 0 to normal inputs (ie acoustic environment, air-conducted and bone-conducted sidetones) and output (voice level), we then assign 'increased level' a value of 1; 'reduced level' a value of -1; and 'further reduced level' a value of -2, and so on. It should be noted that these values are only comparative, and even then only approximately so. For instance, an actual output value of 1 is given to both (a) ear-plugs in quiet, and (b) ear-plugs in noise, simply because we know that voice level in (a) is higher than voice level with open ears in quiet, to which 0 is assigned, but lower than voice level in the Lombard effect, to which 2 is assigned; and similarly, we know that voice level in (b) is lower than in the Lombard effect (2), but (presumably) higher than voice level with open ears in quiet (0). Thus (a) and (b) are assigned the same value, although in fact there is no way of knowing whether these voice levels actually are the same, since none of the reviewed experiments compares them directly. For ear-plugs in noise and ear-muffs in noise, however, we do know from Lower and Martin's (1976) experiment that the output values are not significantly different.

Where the input values are in doubt, which occurs only for bone-conducted sidetone, variants are given, separated by commas; for example, in the case of open ears in noise, it is not known whether bone-conducted sidetone is increased, remains at the same level, or is masked by noise and therefore reduced.

Speaking Condition	INPUTS: Environment	Sidetone: Air-Conducted	Sidetone: Bone-Conducted	INPUT SUM	PREDICTED OUTPUT	ACTUAL OUTPUT	
Open ears in quiet	0	0	0	0	0	0	
Earplugs in quiet	-1	-1	1	-1	1	1	
Open ears in noise	0	-1	(-1),0,1	(-2),1,0	(2),1,0	2	Lombard effect
Earplugs in noise	-1	-1	1	-1	1	1	
Earmuffs in noise	-2	-2	(3)	(-1)	(1)	1	
Increased air-conducted sidetone	-1	1	0,(1)	0,(1)	0,(-1)	-1	Fletcher effect
Increased air-conducted sidetone + noise	0	1	0,(1)	1,(2)	-1,(-2)	-2	Fletcher effect
Reduced air-conducted sidetone*	-1	-1	0	-2	2	1	Sidetone Attenuation effect

Table 1: Input-output analysis of sidetone effects on voice level. (In brackets are variant and hypothetical values giving best fit to actual output.)

* For simplicity we use here Black's (1954) experiment which effectively reduced air-conducted sidetone by inducing temporary threshold shifts in speakers' hearing.

When a row is added, the sum of the row should give the output value (voice level) for that speaking condition when the sum's sign is reversed, if the output compensates for the input.

The output columns give both actual output (as obtained in the reviewed research), and output predicted by the sum of values assigned to the inputs, with the sum's sign reversed.

Comparing actual and predicted outputs enables us to judge whether the assigned inputs are likely to be roughly correct, to decide between variant input values, and to assign a hypothetical input value in cases where we cannot form a guess as to one of the input values in a row. The hypothetical value is arrived at by subtracting the sum of assigned input values from the actual output value. (In fact the Table contains only one hypothetical value - that of bone-conducted sidetone when ear-muffs are worn in noise.)

For the Lombard effect, the best fit is obtained by assuming that bone-conducted sidetone is reduced. The reason that increased bone-conducted sidetone was considered as a variant in the Table is that Dolch and Schubert (1955) found that there is greater sound pressure in the earphones of a speaker's headset than there is directly in front of the mouth, and that this is due to the coupling of a cavity to the vibrating skull. This would suggest that the use of earphones automatically increases the level of bone-conducted sidetone. If this reasoning is applied to the Lombard effect when headphones are worn, the Table no longer predicts the Lombard effect, although it is true that one has no audible sidetone at all under very loud noise masking. If very loud noise masking is accepted as reducing bone-conducted sidetone, then the corresponding value (-1) does predict actual output.

The same problem arises for the Fletcher effect, except that in this case the best fit is given by assuming that bone-conducted sidetone is increased. This assumption is reasonable in the case of increased air-conducted sidetone without admixed noise, but more open to doubt in the case of increased sidetone without noise, although there would be less doubt, as in the case of the Lombard effect, if the noise level were very high. However, the case of increased air-conducted sidetone + noise could equally well be treated by assigning a value of 2 to air-conducted sidetone, and 0 to bone-conducted sidetone, and so the Table does not allow any clear conclusion here.

Attempting to characterise the input/output factors of speech in a table such as this emphasises the interdependence of all the factors involved. For instance, there is a problem as to whether the environment can be made more salient (ie given a + value) or not; that is, does a speaker attend more to the environment at some times than at others? It will be noticed that the Table as it stands includes only 0 or minus values for the environment. Should noise be treated as affecting sidetone only, leaving the environment constant, or should it also be treated as an environmental variable? The point is that, whichever treatment is adopted, at least one of the other values in the row will be affected.

Apart from the Table itself, it must be true empirically that an acoustic change in the environment will affect the character of at least air-conducted sidetone, and that (apart from experimental intervention) a change in sidetone usually betokens a change in the acoustic environment. Furthermore, a speaker aiming to be intelligible is not so much interested in their own voice alone, or in the environment alone, but rather in how their voice interacts with the environment.

What Lane and Tranel are essentially claiming is that the speaker can predict how their voice interacts with the environment from auditory monitoring of the environment alone - in other words, in terms of the Table, only the first and last columns are necessary to characterise the input/output factors in speech. This would be an economic alternative, but it leaves several problems.

Firstly, how do speakers predict from 'environment' to 'environment + voice'? Secondly, this presupposes that, since one's own voice is not treated as part of the environment, it is treated as a special percept, and to treat one's voice as a special percept without making further regulatory use of it seems uneconomical.

In fact, however, Lane and Tranel are forced to bring in what they call 'vocal effort' to fill the regulatory role of the rejected auditory feedback, although they acknowledge this point only in passing. Even so, to attribute voice regulation wholly to vocal effort + acoustic environment + audience feedback, when sidetone is also available, seems perverse. The large role imputed to audience also seems risky, since it then becomes difficult to explain such occurrences as the presence of the Lombard effect when it is obvious to the speaker that the audience is not also in noise, and some public speakers' difficulties in making themselves heard, in spite of audience feedback. Conversely, we know that purely acoustic responses to audience factors are often inappropriate - for example, shouting at foreigners, or those we are angry or impatient with, presumably in the vain attempt to be better understood.

In most situations, audience requirements are correlated with sidetone conditions for the speaker, since both are simultaneously determined by the nature of the acoustic environment. Therefore it is difficult to assign a separate role to audience and to sidetone. Audience requirements and sidetone can, however, be separated experimentally. This is attempted in the next chapter.

CHAPTER THREE

SIDETONE AND VOICE LEVEL

EXPERIMENTS I - VI

The first two experiments to be reported attempt to separate the effect of audience and the effect of sidetone by the strategy of bringing them into conflict, and observing whether one of them over-rides the other in determining voice level.

This approach also offers a way of finding out whether sidetone effects are purely sensory-motor, as the cybernetic view implies, or 'social', as Lane and Tranel claim. If they are sensory-motor, then the sidetone effect should over-ride audience requirements when the two exert pressure in opposite directions, thereby showing that they are truly separate factors, and not two aspects of a single, 'social' factor. If the effects are purely social, then it should be possible to over-ride any effect predicted by the sensory-motor view by giving instructions which demand an opposite effect for audience reasons.

Experiment I: The Effects of Noise on Voice Level in Voice-to-Distance Matching

The speaker's task in this experiment was to produce optimum voice levels for communication with a listener at different distances. The speaker was at the same time subject to different levels of noise within each distance condition. The instructions to speakers stressed that voice level should be gauged accurately to the listener's distance, and that loud shouting would impair intelligibility (it is in fact the case that very loud shouting impairs intelligibility: Pickett, 1956, 1958). It was hypothesised that the Lombard effect, if it is sensory-motor, would conflict with the effect of these instructions, which implied that only one voice level was appropriate for optimum communication at any one distance, and that therefore voice level would rise with increasing noise within a given distance. The null hypothesis, identified with Lane and Tranel's intelligibility-conserving explanation of the Lombard effect, and with their claim that vocal effort is a sufficient cue to autophonic loudness, was that there would be no significant effect of noise on voice level, since the 'audience factor' (distance) should over-ride.

The listeners were instructed as if the experiment were set up to test the speakers' intelligibility, and were asked to write down what they heard on response sheets.

METHOD

Subjects

There were 26 subjects. All either were, or had been, University students, and none had any known speech or hearing disability. 10 men and 3 women were speakers, and their mean age was 25. The listeners comprised 6 men and 7 women, and their mean age was 26.

Apparatus and Materials

White noise was produced by a custom-built white noise generator, which had equal frequency representation between 80Hz and 8000Hz. A cushioned headset was used for binaural presentation of noise to speakers.

A microphone was placed approximately 12" from the speaker, to record voice levels on a Revox tape-recorder for later analysis. A headrest was used to keep speakers' distance from the microphone constant, and thus reduce irrelevant variation in voice levels.

Responses were cued by means of a stop-watch and small signal light.

Different word lists were used for each distance X noise level combination within a speaker, and each list contained 10 words which were selected at random from Peterson and Lehiste's (1960) Revised CNC lists for auditory testing. These are monosyllabic words, phonemically and frequency-balanced.

Procedure

The procedure was carried out in accordance with a 4 (noise levels) by 3 (distances) factorial design, with repeated measures. The roles of speaker and listener were randomly assigned. The distances between speaker and listener were 6½ feet, 28 feet, and 50 feet, and the order of distances was randomised. The noise levels were: no noise, 75dBA, 90dBA, and 100dBA*, the latter three selected as representing subjectively approximately equal intervals of loudness (to the experimenter). The order of noise levels was randomised within each distance for each subject, each noise level and each distance occurring together once.

The speaker was given the following written instructions:-

> "Your task is to correctly transmit sets of words to the listener. The words are typed on the sheets you have in front of you, one set on each sheet. For each set of words, the listener will be at a different distance away from you.
>
> Try to speak clearly enough so that the listener will hear all the words correctly, and at a loudness appropriate for the distance. But <u>do not</u> try to speak at maximum loudness right from the start, for this may simply distort your speech and make it less understandable. Instead, <u>try to gauge your voice accurately to the distance</u>.
>
> Say the first word when the light in front of you flashes. Then wait for the next flash before you say the next word, and so on for each word."

The listener was also given written instructions:-

> "Your task is to write down lists of words, each ten words long, which will be transmitted to you by the speaker.
>
> Sit in one of the three chairs you see spaced out in the hall, according to the number written at the top of your response sheet.
>
> Write down each word as soon as the speaker says it. If you cannot identify a word for

---

* These noise levels were within the intensity X exposure time safety limits specified by Kryter et al (1966).

All dB levels reported are re 0·0002 microbar certain, make a guess. The speaker will not repeat any words.

When you have written down all 10 words from each list, move to the next position marked at the top of your next response sheet."

Intervals of 10 seconds were left between words for the listener to write down responses.

RESULTS AND DISCUSSION

The measurement of voice levels was done from the experimental tapes, using a sound level meter (slow setting). Out of each 10-word list, only the middle 6 words were measured, in order to avoid end effects.

An analysis of variance was carried out on the voice level data, which were in the form of means. (Tables 2 & 3) This showed significant effects of noise ($F = 40.68$, df 3/36, $p < .001$), and of distance ($F = 31.23$, df 2/24, $p < .001$).

The graphed means (Fig. 1) show that voice level increased over distance, but that voice levels within a distance were not constant, but increased with noise, as predicted. Furthermore, strength of association estimates (Kirk, 1968, p 198) indicated that the proportion of the variance accounted for by distance was only .04, whereas the proportion accounted for by noise was .36.

Noise caused these speakers to raise their voice levels, contrary to the instructions they were given, which had stressed that distance should be the sole criterion for voice level, and hence that only one voice level was appropriate for any given distance. The subjects also, then, ignored the information they were given in the instructions to the effect that loud shouting would be likely to impair their intelligibility.

Although these speakers showed the Lombard effect, we must however be impressed by the implications of the systematic nature of the voice level changes observed, which clearly argue against any uncontrolled changes of

Noise Levels	Distances in Feet			
	6½	28	50	$\bar{X}$
No noise	62.05	65.77	69.39	65.74
75 dBA	74.93	78.72	81.23	78.29
90 dBA	76.69	80.12	82.95	79.92
100 dBA	80.06	83.57	85.71	83.11
$\bar{X}$	73.43	77.04	79.82	

Table 2 : Mean Voice Levels (dBA)
Experiment I

Source	Sum of Squares	df	Mean Square	F	P
Noise	6878.92	3	2292.97	40.68	<.001
Distance	804.01	2	402	31.23	<.001
Subjects	7898.52	12	658.2		
Noise x Distance	62.81	6	10.5	<1	
Subjects x Noise	2028.92	36	56.36		
Subjects x Distance	308.83	24	12.87		
Subjects x Distance x Noise	938.35	72	13.03		
Total	18920.36	155			

Table 3: Analysis of Variance on Voice Level Means Experiment I

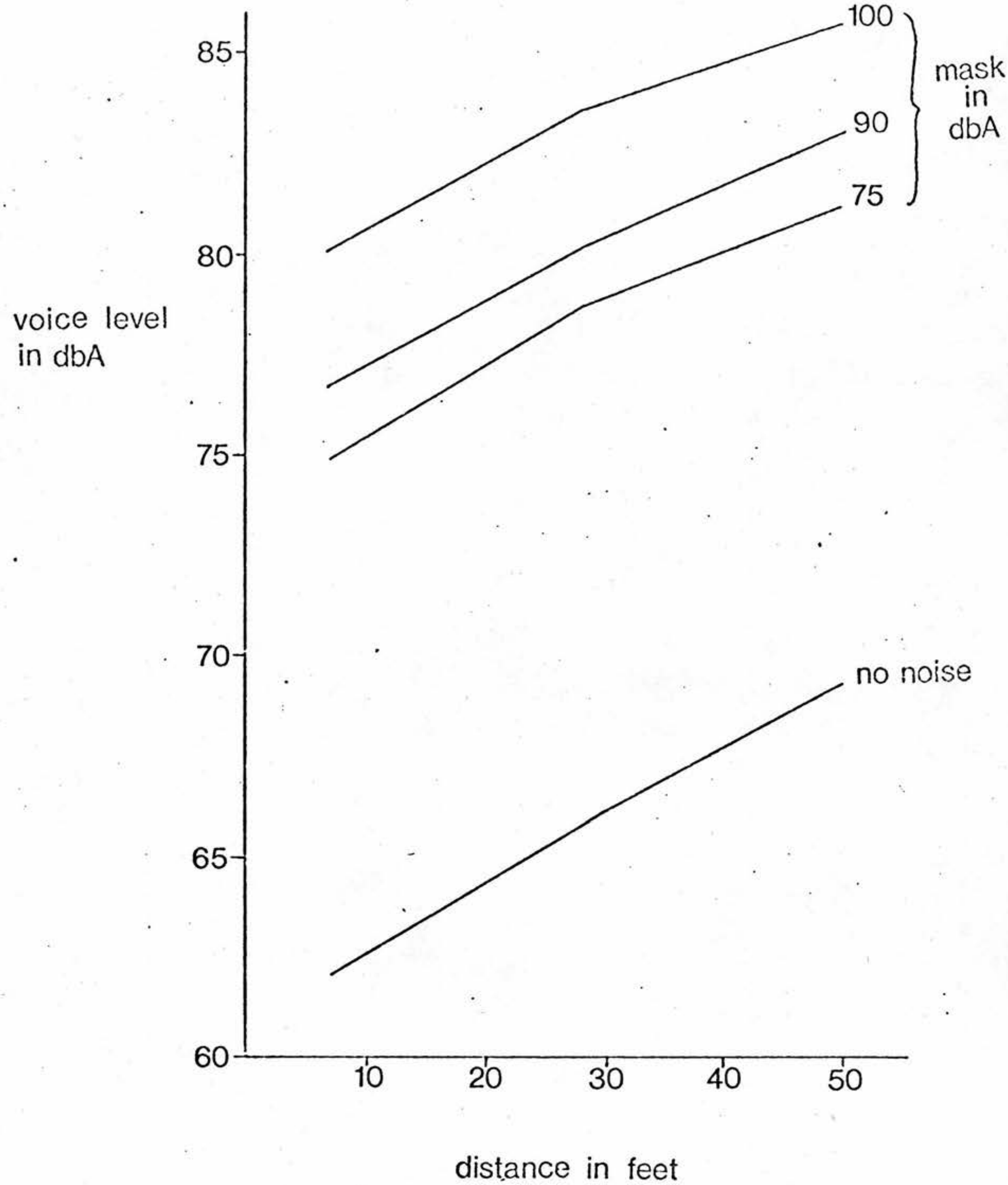

Figure 1: Mean voice levels varying with masking noise level and distance of listener, Experiment I

voice level in speakers' attempts to carry out the experimental task. Voice levels increased systematically with noise within a distance, and with distance within a noise level (Fig. I).

While it is true that speakers did not stick to a single voice level for one distance, as they were instructed to do, they perhaps showed superior wisdom by responding in a manner which appears functional if noise really were a factor threatening their intelligibility to the listener. If the environment gets noisier, one has to shout louder in order to be understood, and this applies for any fixed distance also.

The main objective of this experiment was to compare speakers' voice-to-distance matching performance with, and without, masked self-hearing. Although it is clear that performance with masking was not impaired relative to performance without masking (contrary to what Lane and Tranel's theory would predict) it should be pointed out that these speakers' performance <u>without</u> masking did not compensate fully for distance, nor did their performance <u>with</u> masking compensate fully for noise.

Regarding distance compensation, Beranek (1954)* has presented data showing how received voice level (i.e. received by the listener) falls off with distance from a speaker who is not attempting to compensate. For perfect distance compensation, the speaker should increase voice level so as to cancel out the fall-off found by Beranek. For a distance of 26' Beranek found a fall-off of 11dB, whereas the present experiment found a compensation of less than 5dB.

For noise compensation, however, the levels reported in the present experiment are much less (i.e. only about 4dB in voice level for 25dB of noise). This is congruent with the fact, as Lane and Tranel point out, that voice level compensation for noise varies with the nature of the speech task, and is much less when word lists are read out, as in the present experiment, than when the communication is authentic.

---

*Cited in K.D. Kryter THE EFFECTS OF NOISE ON MAN
Academic Press, London, 1970

We are left, then, with a discrepancy between the distance compensation and the noise compensation in this experiment, the latter being poorer.* It is possible that the emphasis on accurate distance compensation in the instructions led to better performance in this case.

However, it is important to point out that, while speakers were only mildly flexible in their response to environmental factors, they are much less flexible in their response to the audience factors which Lane and Tranel make the central feature of their theory. The speakers in this experiment made <u>some</u> response to their acoustic environment, but they totally failed to take account of the fact that their listeners were evidently <u>not</u> in the same acoustic environment as themselves. It was plain for them to see that their listeners were not exposed to the noise, and in addition many of the listeners made no attempt to conceal their astonishment at the extremely high voice levels generated by the speakers, particularly at the shortest distance (6½ feet). To Lane and Tranel's intelligibility-conserving theory of the Lombard effect must, then, be added the rider that speakers always behave as though their listeners are in the same acoustic environment as themselves, even when the evidence before them is to the contrary. In view of Pickett's (1956, 1958) finding that loud shouting can impair intelligibility, it seems unlikely that speakers would remain intelligible in high levels of noise, even if their listeners were exposed to the same noise.

Experiment II: The Effects of Ear-Plugs on Voice Level in Voice-to-Distance Matching

INTRODUCTION

Of three previous studies of the effect of ear-plugs on voice level, one found an increase (Kryter, 1946), one found a decrease (Hebb, Heath and Stuart, 1954), and one found that the effect varied between a decrease and no effect at all (Garde, 1964).

Lane and Tranel (1971) considered these studies in their review, and decided in favour of a decrease in voice level, since this seemed explicable in the following terms. Ear-plugs attenuate the air-conducted sidetone of one's voice, consequently the bone-conducted sidetone will sound relatively louder than normal. As a compensation for this apparently louder-than-normal sidetone, the speaker reduces voice level. In accordance with their intelligibility-conserving theory, however, Lane and Tranel regard this as an adjustment to the listener, rather than to sidetone per se. This interpretation is independent of their reasoning regarding the likely effect of ear-plugs on voice level, which seems straightforward.

The point of this experiment, designed as a complement toExperiment I, was again to bring audience requirements into conflict with a sidetone effect, this time using ear-plugs in an attempt to reduce speakers' voice levels inappropriately for the task of communicating with a listener over distance. It was hypothesised, then, that ear-plugs would induce under-shoot of appropriate voice level for a given distance, 'appropriate voice level' being the voice level produced for that distance in the control, or no ear-plugs, condition.

METHOD

Subjects

Speakers: these were the subjects who had been listeners in the previous experiment.[1] 7 of the speakers were men, and 7 were women. Their mean age was 25.

Listeners: These subjects had acted as speakers in the previous experiment.[2] 11 of the listeners were men, 3 women, and their mean age was 25.

Apparatus and Materials

The ear-plugs used in this experiment were malleable wax 'mufflers', obtained from Boot's chemists.

Responses were cued and recorded as in the previous experiment, and a different set of words lists was randomly selected from the Peterson and Lehiste (1962) lists, as before, resulting in completely new lists.

Procedure

At the end of the previous experiment, subjects were informed that they would now exchange roles, and were given the same written instructions for speaker and listener as in the previous experiment. The procedure henceforth was as before. The order of conditions (ear-plug condition and no-ear-plug condition) was alternated between subjects, and each distance occurred once, in random order, within each ear-plug condition.

---

1. The apparently extra subject appearing here is explained by the fact that in the previous experiment the data from one pair of subjects had to be discarded because of excessive extraneous noise during the experimental session.

2. Again there is apparently one extra subject, for the reason given above.

RESULTS AND DISCUSSION

The data were analysed as in the previous experiment. A two-way factorial analysis of variance for repeated measures was carried out on the voice level data. This showed significant effects of ear-plugs ($F = 5.13$, $df = 1/65$, $p < .01$), and of distance ($F = 65.75$, $df = 2/65$, $p < .001$) (Tables 4 and 5). Figures 2 and 3 show that the effect of ear-plugs was in the opposite direction to that predicted, ie ear-plugs increased voice level, and that voice level increased systematically over distance.

The results of this experiment give rise to much the same conclusion as the previous experiment. Although ear-plugs, like noise, did not impair the ability to assign relative vocal magnitudes to distances, the control of absolute voice level, which was stressed in the instructions, was either impaired for sensory-motor reasons, or traded for social reasons, speakers assuming that listeners were in the same acoustic conditions as themselves.

The finding that ear-plugs significantly increased voice level is in agreement with Kryter's (1946) finding, although the effect found here, being of the order of only 1dB, is smaller than that found by Kryter, which was 3dB. Of the two studies which had different findings, the study by Hebb et al. (1954) required subjects to wear ear-plugs over a period of several days, and Garde (1964) carried out an informal experiment in which he used only his subjective judgment in determining voice level. The adequacy of this procedure may be judged in the light of the present experiment in which the effect was not apparent at all to the 'naked ear', but was only discovered from the sound level measurements. Indeed, one's subjective judgment would have been in agreement with Garde. It is possible, however, that the speakers in this experiment attempted to keep their voices down since, as listeners in Experiment I, they had heard some surprisingly high, and inappropriate (to them) voice levels.

This slight ear-plugs effect would seem unlikely to enhance intelligibility, contrary to Lane and Tranel's view. However, supposing that the speaker behaves as though the

	Distances in Feet			
	6½	28	50	$\bar{X}$
No Ear-plugs	62.22	66.01	69.14	65.79
Ear-plugs	62.99	67.39	70.49	66.96
$\bar{X}$	62.61	66.70	69.81	

Table 4: Mean Voice Levels (dBA)
Experiment II

Source	Sum of Squares	df	Mean Square	F	P
Total	5639.93	83			
Distance	732.47	2	366.23	65.75	<.001
Ear-plugs	28.58	1	28.58	5.13	<.01
Distance x Ear-plugs	1.7	2	0.85	0.15	
Subjects	4516.66	13			
Pooled Error	362.22	65	5.57		

Table 5: Analysis of Variance on Voice Level Means Experiment II

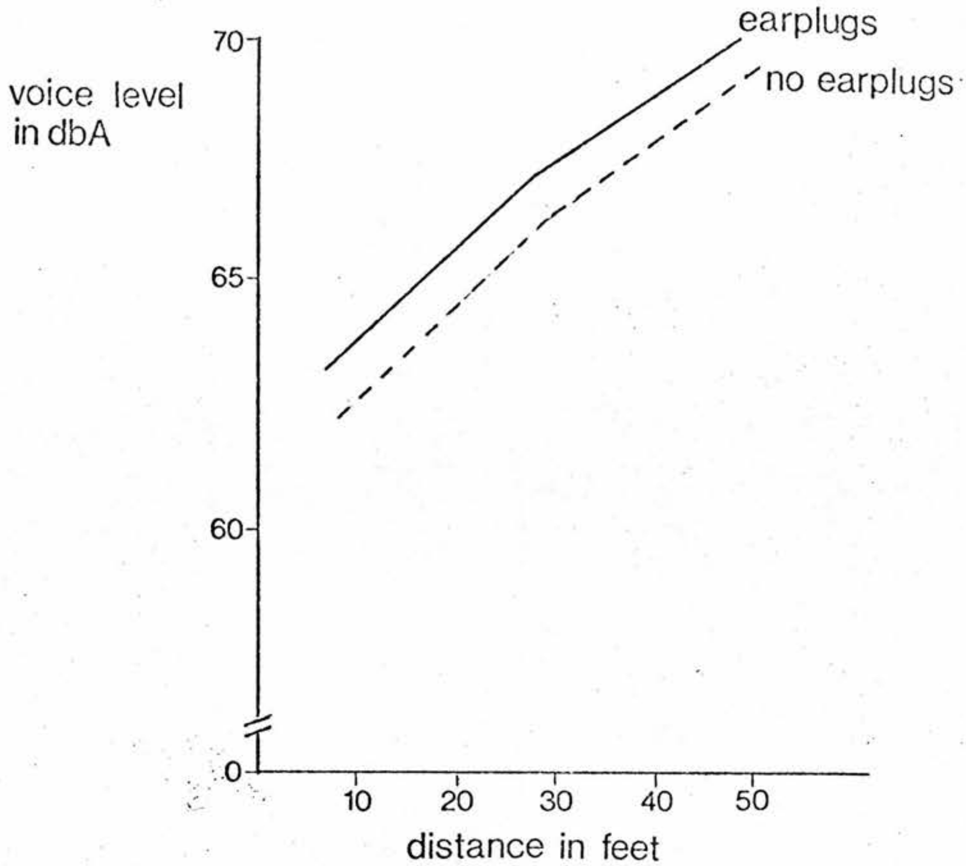

Figure 2: The effect of earplugs on voice level Experiment II

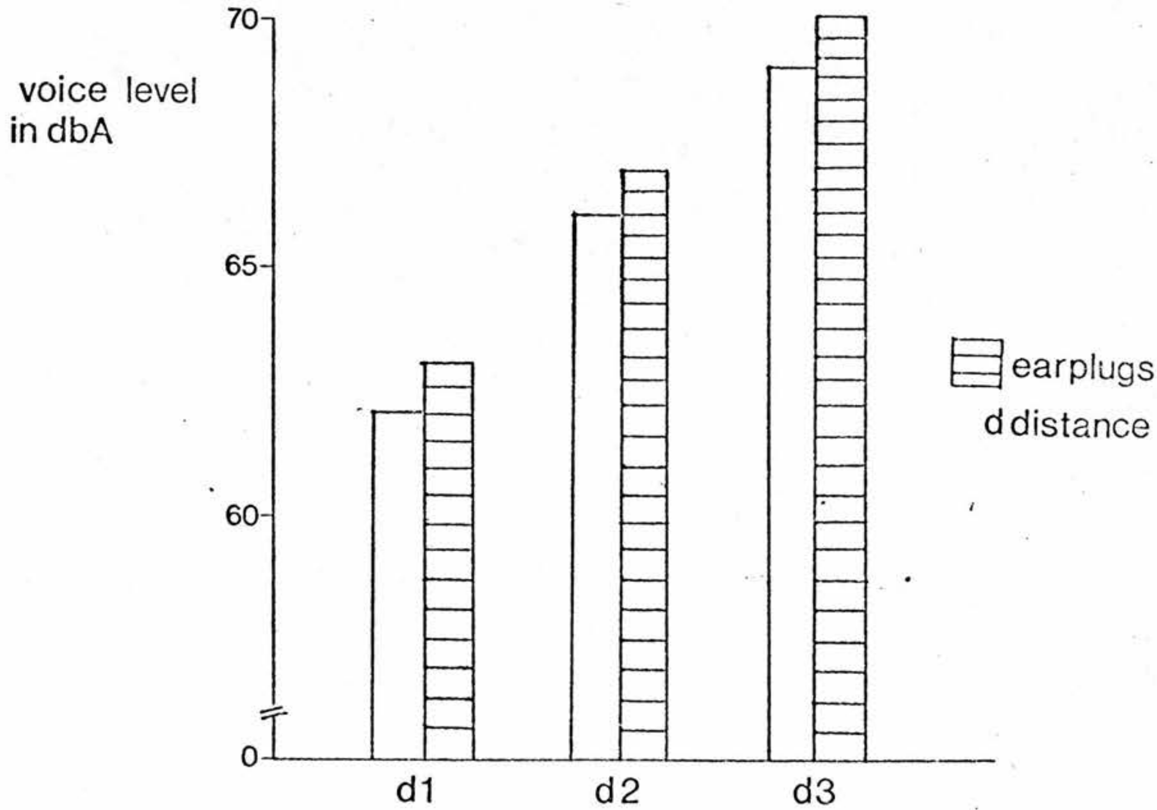

Figure 3: The effect of earplugs on voice level Experiment II

listener were also wearing ear-plugs, the intelligibility of 'ear-plugged' speech should strictly be tested with 'ear-plugged' listeners.

Finally, the finding by Hebb et al.that voice level dropped after their subjects had been wearing ear-plugs for some days does argue in favour of a long-term sensory-motor effect, comparable to the difficulties that some adventitiously deaf people have in controlling their voice level (Espir and Rose, 1970). It certainly seems to argue against a social interpretation, since it involved lowered intelligibility.

Overall, then, the first two experiments indicate that the speakers were certainly affected by audience considerations, but that their attempts to accommodate to an audience were constrained by an involuntary strategy which, in the first experiment at least, fired broad-shot at what was presented to them as a very narrow target. To mix the metaphors, they seemed to switch on to an 'automatic speaker' who appeared to have the sole aim of being heard, regardless of all other considerations.

This strategy, we must suppose, was in the service of communication. The question then arises - what would happen if subjects were vocalising, but not communicating, in similar conditions? If this strategy is truly a communicative one, we would expect it not to occur, or at least not with such force as in the first experiment. Conversely, if such 'sidetone effects' are predominantly under sensory control, we would expect them to occur to a similar degree in both communicative and non-communicative vocalisation.

Experiment III tests these hypotheses by investigating the effects of sidetone manipulation on non-linguistic vocalisation, a sustained vowel. White noise masking and ear-plugs were used to manipulate the speaker's acoustic environment, as in the two preceding experiments.

In addition, it was hypothesised that any kind of actual or expected sensation on the ear might produce a Lombard effect and so a condition was included in which dummy electrodes were attached to speakers' ears.

Experiment III: The Lombard Effect in Non-Linguistic Vocalisation

METHOD

Subjects

Undergraduate, postgraduate, or former students served as subjects. 9 were men, and 9 women. Their mean age was 25. None had any known speech or hearing disorder.

Apparatus

Masking noise of 75, 90, or 100dB was produced by a Department-built white noise generator whose frequency output was level between 80 and 8000 Hz. This noise was delivered binaurally to subjects through a cushioned head-set.

The dummy electrodes were large self-adhesive electrodes, wired up but attached with sellotape only to the back of an audiometer, which lit up when 'switched on' by the experimenter.

The ear-plugs were of the same type as those used in Experiment II.

A small signal light in front of the subject was used to cue responses, which were timed with a stop-watch.

Responses were recorded for later analysis, with the microphone placed approximately twelve inches away from the speaker.

Procedure

Subjects were given the verbal instruction that they would be asked to produce a sustained vowel, 'a'*, for 10 seconds under six different hearing conditions. The vowel should be started when the cueing light came on, and continued until it went off. The attempt should be made to keep the voice 'the same' throughout all conditions in the experiment.

The subject produced one sustained vowel under each

---

* Since it was sustained for 10 seconds, this vowel was not equivalent to the utterance 'ah!'

of the following 6 sidetone conditions:- with 75, 90, or 100 dBA masking noise, with dummy electrodes on the ears, with ear-plugs in the ears, or with open ears. These conditions were presented in a different random order for each subject.

RESULTS AND DISCUSSION

The peak amplitude of each sustained vowel was measured on a sound level meter (slow setting). There was a clear peak in each vowel, usually at the beginning of the vocalisation.

An analysis of variance for repeated measures showed a significant experimental effect ($F = 53.18$, $df = 5/85$, $p < .001$). See Tables 6 & 7. Scheffé multiple comparisons showed no significant differences within non-masking conditions (control, dummy electrodes, and ear-plugs) or within the masking conditions (75, 90, and 100 dBA). However, there was a significant difference between the dummy electrode condition (selected as having the highest sum within the non-masking conditions), and the combined masking conditions ($F = 119.78$, $F' (\alpha < .001) = 23.8$, $df = 5/102$, $p < .001$).

This experiment has, then, demonstrated a clear Lombard effect in non-linguistic vocalisation, arguing against the exclusively intelligibility-conserving interpretation of sidetone effects by Lane and Tranel.

However, it is interesting to note that the pattern of effects is somewhat different to that obtained with linguistic vocalisation in the first two experiments. Firstly, there was no significant effect of ear-plugs this time. Secondly, subjects hit a voice level ceiling with the 90 dBA mask, though this is explained by the fact that the subjects had much higher voice levels in the control condition in this experiment as compared to the preceding two experiments (see Fig. 5)*. Some such ceiling factor may also have operated to obscure any effect of ear-plugs in this experiment; Fig. 5 shows that the control voice

---

*Possibly because of the very high energy contained in a sustained vowel.

	Ear Plugs	Control	Dummy Electrodes	75dBA Mask	100dBA Mask	90dBA Mask
Voice Level in dBA	80.06	80.11	80.83	91.22	94.06	94.25

Table 6: Means of Peak Amplitudes of Sustained Vowels, in Increasing Order of Magnitude Experiment III

Source	Sum of Squares	df	Mean Square	F	p
Treatments	4563.34	5	912.67	53.18	<.001
Subjects	5746.21	17	338.01		
Treatments x Subjects	1459.20	85	17.16		

Table 7: Analysis of Variance on Mean Peak Amplitudes of Sustained Vowels Experiment III

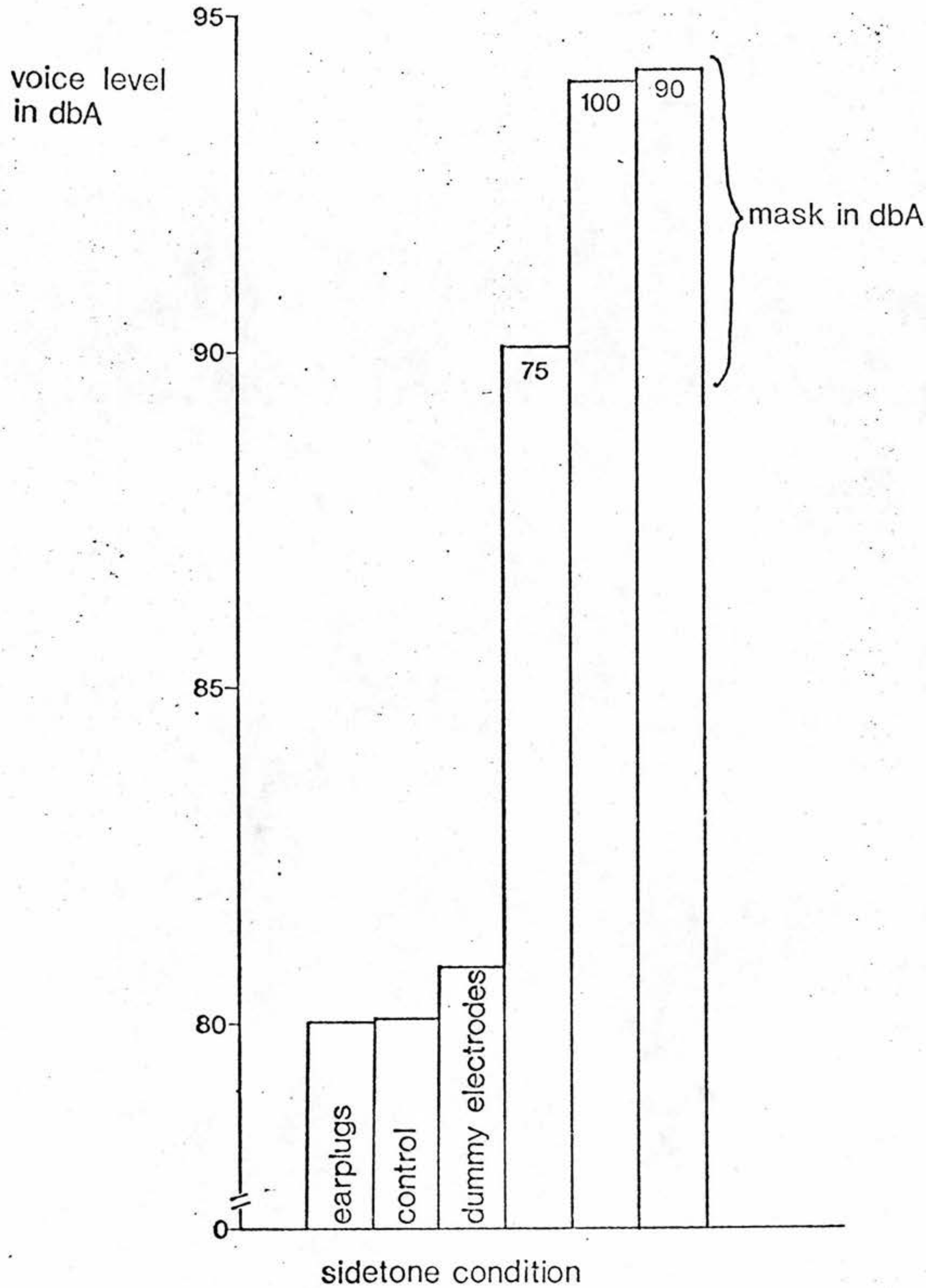

Figure 4: Voice Level in different sidetone conditions, Experiment III

level mean was already far higher than the highest ear-plugged voice level mean in Experiment II. In other words, it is possible that the ear-plug effect has a specific voice-level range of occurrence, and this may be one reason for the conflicting results in previous experiments with ear-plugs (see discussion on p.44).

The absence of any effect of the dummy electrodes precludes any easy inferences about 'psychological' or expected effects, or even irrelevant sensory effects; clearly none were created here.

In conclusion, then, the Lombard effect is not confined to communicative situations where the conservation of intelligibility is a motivating factor; it also occurs, and just as strongly, in a task as purely sensory-motor as it seems possible to devise for the human voice. This conclusion must, however, be qualified by just this consideration, that no vocalization may be *totally* non-communicative. In this case, however, Lane and Tranel's thesis that sidetone effects ~~a~~ of voice are primarily communicative, rather than sensory-motor, must be replaced by the somewhat diluted proposition that, since *any* vocalization may be a form of communication, an intelligibility-conserving motive could never be ruled out. Rather, we are left with a scale on which, as they indeed point out, sidetone effects increase in strength with the communicative nature of the experimental task, but this is a scale which has no zero point.

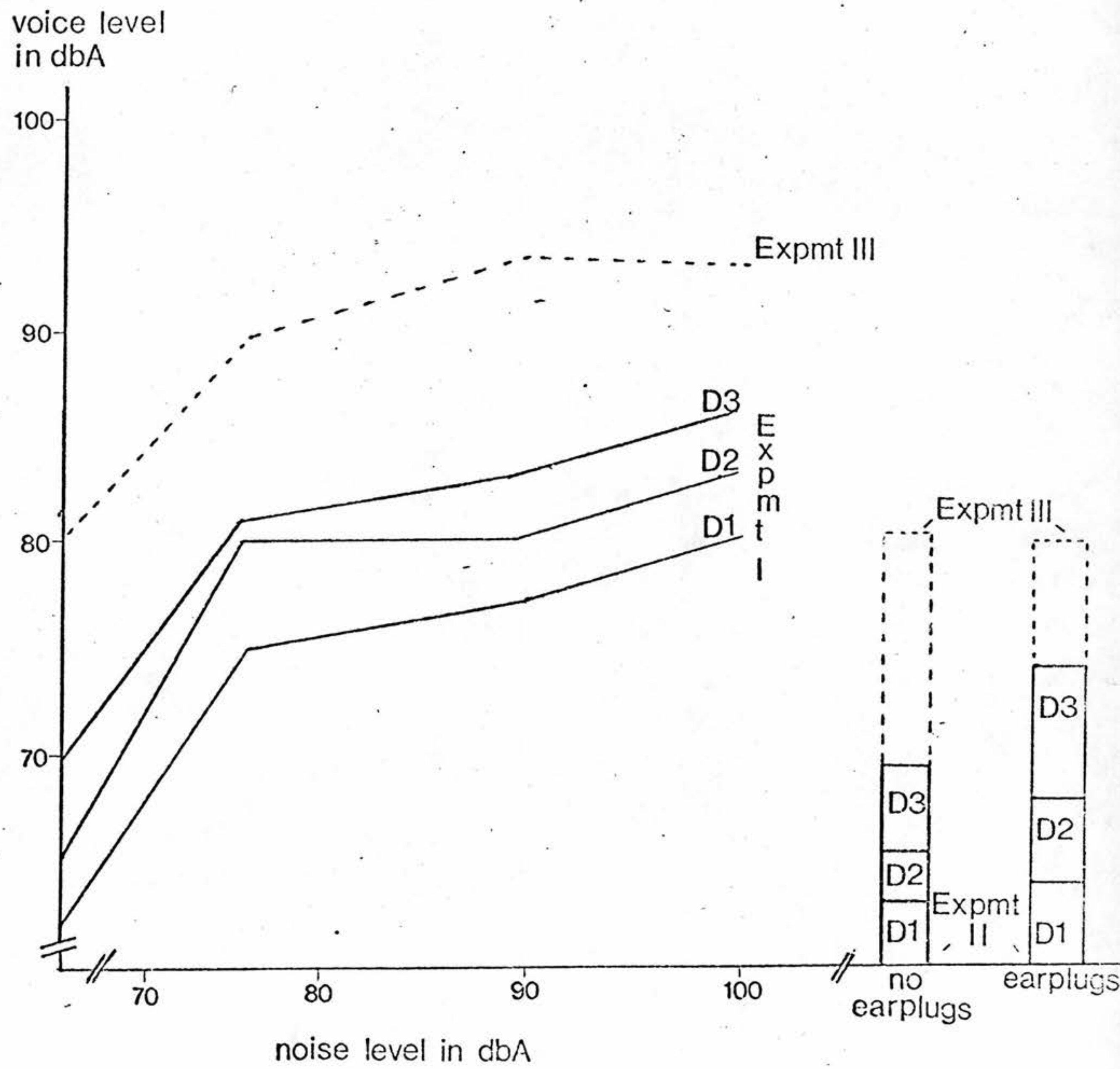

Figure 5: Comparison of mean voice levels in Experiments I, II, and III (D = Distance)

Experiment IV: A Right-Ear Superiority in the Lombard Effect

Having demonstrated in Experiment III a Lombard effect in non-linguistic vocalization, the question next arises - how far is the Lombard effect related to the activity of communicating when the two occur together?

We already know from Lane and Tranel's review that the effect is stronger when there is a listener in the situation, but we also know that the effect is involuntary, and can occur outside the speaker's awareness.
The fact that the ear preferences are different in the psychological processing of speech sounds and of environmental sounds (Bakker, 1967, 1968) offers a strategy for locating the psychological level at which noise comes to be treated as relevant to communication by the speaker. In other words, it offers a strategy for determining whether noise is treated by the speaker as being part of the environment, or as a factor integrally involved in the activity of communicating.

Since there is some evidence that the left hemisphere, in most right-handed subjects, is the one which dominates language functions, (Boone, 1959; Curry, 1967) masking of the right ear might be expected to produce a greater Lombard effect if, as Lane and Tranel argue, the regulation of voice level is based on an assessment of acoustic conditions in so far as they relate to communication, rather than merely to the adjustment of speech as motor output. If, on the other hand, noise heard during speech were perceived as purely environmental noise, with implications at the level of motor performance but not at the level of communication, then we would expect a left ear superiority for the Lombard effect, since specifically environmental sound such as traffic noise, applause, hammering, etc. (Lefevre et al., 1977) is processed preferentially when received through the left ear.

Our expectation that right-ear masking will be more effective in producing the Lombard effect is strengthened by the consideration that one's own voice is probably monitored integrally with, or even as an index of, the acoustic conditions for communication (see pp 45-49 for this argument). In this case, also, we would expect the left hemisphere to be the one which monitors the single, but composite percept 'voice + acoustic conditions', more effectively.

In effect, what is being proposed here is that noise can be monitored preferentially either through the left ear, or the right ear, depending on whether one may be called upon to speak simultaneously with the noise, and therefore on whether the noise has communicative relevance or not.

There is additional reason to expect greater effectiveness of right-ear masking in producing the Lombard effect, in that two studies have found a right-ear advantage for sound contingent on the speaker's articulations. Abbs and Smith (1970) found that speech DAF was more disruptive in the right ear, and Sussman and MacNeilage (1975) found a right-ear advantage when a tone in the right ear was to be controlled with a speech articulator (tongue or lower jaw) to match a pure tone of varying frequency in the other ear. There was no such right-ear advantage if the cursor tone was hand-controlled. This finding led Sussman and MacNeilage to conclude that there is a speech-related auditory sensory-motor integration mechanism in the left hemisphere.

The experimental hypothesis was, then, that right-ear masking would raise voice level higher than left-ear masking. This hypothesis was cast into the following more specific prediction: when voice level without masking (control condition) was subtracted from voice level with monaural masking, then the score for right-ear masking minus the control score should be greater than the score for left-ear masking minus the control score, thus:

(right-ear masked minus control) > (left-ear masked minus control)

It should be noted that there is no intention here to infer the causes of asymmetry effects per se. Whether these are due to the physiological structure of the hemispheres, or non-structural factors such as selective attention, is still largely unknown. For the present hypothesis, that noise during speech is treated as communicatively, rather than environmentally, relevant, it need only be demonstrated that the noise is, like speech, but unlike environmental sound, preferentially processed when received through the right ear. Thus this hypothesis carries no assumptions as to the further nature of the processing beyond ear preference.

METHOD

Subjects

There were 26 subjects, of whom 12 were men, and 14 women. All were either undergraduates or postgraduates, or had had some tertiary education. None had any known speech or hearing disorder, and all were right-handed according to self-report.

Apparatus and Materials

A white noise of 100dBA was generated using the apparatus described in Experiment I, and this was fed to one ear of the subject's headset. A microphone placed roughly 12" from the subject recorded responses onto a Revox tape-recorder for later analysis.

Lists of 10 words for oral reading were constructed from Peterson and Lehiste's (1960) Revised CNC Lists for auditory testing.

A sound level meter (slow setting) was used to measure voice levels.

Procedure

Most of the data were obtained during the course of class demonstrations to undergraduates of sidetone effects. This at least had been the intention, but since fewer data than had been anticipated could be obtained in this way, the number of subjects had to be supplemented by calling on postgraduate students and other volunteers. For this reason, the conditions of the experiment were not the same for all subjects - some were before an audience of classmates and others were not, for instance.

Furthermore, the actual conduct of procedure for each subject was not as formal as in the other experiments reported. The subject was simply presented with a typed list of 10 words, and asked to read out the list once with, and once without, noise masking. The order of masking and no-masking conditions was balanced across subjects. The decision whether to mask the left or right ear of each subject was taken in accordance with the requirement to obtain equal numbers of masked left and masked right ears.*

RESULTS AND DISCUSSION

The middle six out of the ten response words for each subject were measured, to avoid end effects. Two sets of means were obtained: (i) the mean difference between voice levels in the no-masking condition, and voice levels in the masked right ear condition, and (ii) the mean difference between voice levels in the no-masking condition, and voice levels in the masked left ear condition. The overall means were:

Right Ear Masked	Left Ear Masked
8.86	6.96

The difference between the means is in the predicted direction, and is significant by a one-tailed test: $t = 1.81$, $df = 24$, one tailed $p < .05$.

This result suggests that the right ear is indeed preferentially involved in the regulation of voice level, and therefore that this regulation may be as much

---

* The collection of data for one masked ear only from each S was an uneconomical design in the event, but would not have been so had the large number of subjects anticipated materialised. If they had, the one-ear-only measure would have been justified on the grounds of its avoidance of carry-over and order problems, and its brevity for the S.

an integrally communicative as a sensory-motor process. It should be noted though that this interpretation depends on acceptance of two premises: that hemisphere specialisation for language exists, and that it was invoked in this study. It is possible that, since there is cross-talk between ears with high masking levels (Zwislocki, 1953), raised voice level with primary masking on either ear could simply reflect selective attention to the primary-masked ear. As this study found such a small ear difference, such an alternative explanation cannot be definitely rejected.

Subject to these qualifications, it may, then, be suggested that when the Lombard effect occurs with linguistic vocalisation, noise is processed more effectively when received in the right ear, and hence appears to be processed in a manner similar to speech, rather than non-speech stimuli. This would support Lane and Tranel's contention that noise producing the Lombard effect is treated by speakers as immediately relevant to communication, rather than to sensory-motor control of movement.

Considering the main points of the preceding experiments in the light of the sensory-motor and social theories, how does each theory fare so far?

The sensory-motor theory's interpretation of the first two experiments is that speakers were unable to set appropriate voice levels in communicating with a listener at a distance because the sidetone manipulations interfered with their ability to hear themselves, and so set correct values for voice level. This interpretation faces the problem that changes in subjects' voice levels were never random, but always systematic.

The social theory's interpretation is that, in both experiments, speakers' overriding concern was to respond to audience requirements. This interpretation fails to explain, in Experiment I, speakers' apparent insensitivity to the obvious fact that their audience could not hear the same noise as themselves, as well as to cues from the listeners that they were speaking too loud.

It is possible, however to combine these two theories' explanations of the results in such a way that the objections to each explanation alone are overcome. Thus, in line with the sensory-motor theory, it may be the case that sidetone interference prevents voice regulation, but that auditory voice regulation is not needed for setting relative voice level* (only for setting absolute voice level). The social theory then explains why voice levels changed systematically with sidetone interference - ie to conserve intelligibility (the sidetone manipulation being treated by the speaker as merely a change in the acoustic environment). With this combination of the two theories, the Lombard effect appears as a rather crude intelligibility-conserving strategy. It is a rather 'blind' raising of voice level with noise - not so blind that the voice level increase is not systematically related to the noise increase, yet blind enough to fail to set the voice level appropriately for the audience conditions which clearly obtain (ie listeners not in noise, and who show surprise at the high voice levels used). The crudeness of such a strategy may be compared with that of continuing to use visual signals such as facial expression, posture, and gesture, which are ill-adapted to the listener, when talking on the telephone.

When seen in this way, the Lombard effect ties in well with Lane, Tranel and Sisson's explanation of the Lombard, Fletcher, and sidetone attenuation effects as 'half-way' compensations of voice-level for changed acoustic conditions (see p 43). Indeed, they themselves go on to account for the fact that the compensations are only half-way by suggesting that the speaker is "misled by the disparity in the sensory operating characteristics of speaking and listening" (p 618). Put perhaps more accurately, this explanation is that the speaker is misled by the change in the sensory operating characteristics of sidetone, a situation which prevents a fully adaptive

---

* The fact that one can produce short vocalisations which are very loud, or very soft, from the moment of voice onset shows that one does not need self-hearing in order to do this.

response to the changed acoustic conditions.

Supposing, however, that sidetone were a completely accurate guide to one's speech performance in normal circumstances; it could not continue to be so if the acoustic environment then changed in any way, because if the speaker then compensates for the change, they have no way of knowing, on the basis of sidetone, whether they have compensated successfully, or even at all, since the compensated speech itself is also filtered, as it were, through the changed acoustic environment. So this new, compensated performance must be carried out without accurate feedback. This may account for the exaggerated character of speech under sidetone effects such as the Lombard effect and the DAF effect. They are analogous to groping in the dark, or drunken walking - not mere breakdowns in performance but crude, yet systematic or stereotyped responses which may not be very effective, but represent the best one can do in the circumstances. The question then arises as to why these stereotyped responses should occur at all, in view of their apparently minimal adaptedness. An answer is suggested by analogy with the theory of Natural Selection, according to which even slight adaptedness will confer advantage on individuals or species. In trying to account for the existence of pain, for instance, Darwin writes

> "... pain or suffering of any kind, if long continued, causes depression and lessens the power of action; yet is well adapted to make a creature guard itself against any great or sudden evil ... suffering is quite compatible with the belief in Natural Selection, which is not perfect in its action, but tends only to render each species as successful as possible in the battle for life with other species ..." (Barlow, 1958, pp 89-90; emphasis added)

Similarly, it can be argued that stereotyped responses are in fact efficient in so far as they merely raise the probability of a successful outcome.

The first two experiments, then, support a combined sensory-motor and social interpretation of the Lombard effect, according to which absolute voice level cannot be gauged without normal sidetone, but this situation is compensated for by an intelligibility-conserving strategy of a crude kind which 'blindly' raises voice level with the level of acoustic interference.

Experiment III shows that the Lombard effect also occurs in non-linguistic vocalisation, and even more strongly (in terms of the high voice levels produced, if not clearly in terms of slope). The effect is not, therefore, exclusively linked with communicative situations, and so we must acknowledge that there is a component in the Lombard effect, presumably a pure sensory-motor component, over and above (or perhaps, rather, 'beneath') any intelligibility-conserving component. While it is possible to argue that when subjects vocalise non-linguistically they merely generalise from their linguistic vocal experience and behaviour, the impossibility of testing such a notion forces us to discard it (even though it would be consistent with the 'crude strategy' view of the Lombard effect which we have been putting forward).

Experiment IV attacks more directly the question of whether, and how far, the Lombard effect is language-related, and provides some evidence that it is so related, at least in so far as the noise is preferentially processed if, like language stimuli, it is received through the right ear.

These experiments raise a number of questions which should be pursued, and clearly there are several directions which the thesis could have taken from this point. Why were the voice levels so much higher in non-linguistic vocalisations (Experiment III) than in linguistic vocalisation (Experiments I and II), with the same levels of masking noise? Experiment IV stands in need of replication with either a different design or a larger number of subjects. It also suggests that there would be a left-ear superiority in the Lombard effect with non-linguistic vocalisation.

However, instead of pursuing the further questions raised by these first experiments, the decision was made to persevere in seeking an answer to the question which all of these experiments approach from different directions, namely: do sidetone effects in speech reveal the operation of a sensory-motor, auditory-vocal control system, or do they reveal the workings of a system of stereotyped responses which function to preserve the intelligibility of communication under adverse acoustic conditions? Perhaps it should be reiterated here that these can be substantively different interpretations of the same phenomena, as Lane and Tranel (1971) would claim. It is possible, for instance, that the Lombard effect is a response to noise per se, as an obstacle to communication, rather than to noise as a masker of 'auditory feedback', and that voice level regulation is accomplished entirely through vocal effort plus environmental monitoring. On the other hand, it is possible, as we have pointed out earlier (p 84) to account for more of the observed facts by combining these two interpretations. The sensory-motor interpretation explains the inability to finely gauge voice level to the socio-acoustic setting, but the intelligibility-conserving interpretation explains the systematic core of the stereotyped response.

It was decided, then, to address the concluding experiment in this series of experiments on sidetone and voice level to a test of this combined theory.

Experiment V The Effects of Noise Masking on Voice Loudness Matching and Autophonic Judgment

The combined theory we have just put forward claims that sidetone is not needed when making relative changes of voice level in accordance with changed acoustic conditions, and indeed that these changes will occur quite automatically and involuntarily. At the same time, the theory claims that sidetone is needed in order to produce, or to judge, any specified absolute voice level.

Experiment VI was designed, then, to determine (i) the extent to which subjects are able to keep a constant voice level under different sidetone conditions, by asking them to match a recorded standard voice level, and (ii) how accurately they are able to judge their absolute voice level under different sidetone conditions.

These questions also relate to the wider issue, discussed earlier (pp 29-34) as to whether sensory inflow is needed for the fine control of movement. If subjects cannot match an external standard under different side-tone conditions, this may be because they cannot judge their own voice, in which case we would expect both poor matching performance and inaccurate judgments. On the other hand, they may not be able to match the standard because, although sidetone is irrelevant to the motor control of their performance, some drive to preserve intelligibility forces them to raise their voice level contrary to the instruction to match, even though the task is a non-communicative one. There is a possibility that, in this case, their judgments of their own voice level are nevertheless accurate, and indeed their judgments should be accurate if, as Lane and Tranel claim, speakers use only vocal effort in judging their own voice level, and not sidetone.

The experimental task was one of loudness matching and magnitude estimation. Lane, Catania and Stevens (1960) reported an experiment involving a similar task, in which subjects were asked to produce vocalisations which centred the needle on the dial of a sound level meter (SLM) whose gain was varied by the experimenter (thus varying the sound level required to centre the needle each time), and to assign loudness values to the vocalisations thus produced. Since their subjects were able to assign correct relative values to their vocal productions within different sidetone conditions (open ears, headphone, noise), Lane et al concluded that subjects judged the loudness of their vocalisations in terms of vocal effort, and not using sidetone. However, Lane et al were not interested in the absolute accuracy of their subjects' judgments, and it is clear that the possibility remains that sidetone is needed for such absolute judgments.

In the present experiment, an autophonic scale was first set up for each subject, under a relatively 'normal' sidetone condition (wearing a headset). The autophonic scale consisted of a set of voice levels produced by the subject (objectively measured in decibels), together with the subject's numerical estimates of the subjective loudness of these voice levels. The procedure used was similar to that of Lane et al. The subject produced 'ah' at different loudnesses in order to center the needle of a SLM whose gain was varied by the experimenter, and assigned loudness values to these vocalisations. However, the procedure differed from Lane et al's in that the judgments were made with reference to a fixed, external standard. This was a recorded male-voice 'ah' at medium loudness, and it was given the arbitrary loudness value 10.

In the second part of the experiment, subjects simply attempted to match this same standard, but under different sidetone conditions (headphones, and two levels of white noise - 65dBA and 75dBA). The subjects then assigned magnitudes (again with reference to the standard) to their matches, indicating either a perfect match (10), or some other magnitude representing under- or over-shoot.

From the autophonic scales initially set up for each subject, it was possible to predict, by means of a regression procedure, each subject's expected magnitude estimation for any given vocalisation. It was hypothesised that the sidetone manipulation of the second part of the experiment would impair absolute judgments of vocal magnitude when compared with the judgments in the first part of the experiment (autophonic scaling). In other words, it was hypothesised that the difference between the regression-predicted magnitude appropriate to any particular vocalisation, and the actual magnitude assigned, would increase as sidetone cues were decreased by white noise.

METHOD

Subjects

There were 24 subjects, all of whom were either undergraduate or postgraduate students. The sexes were equally represented, and the mean age was 22. None had any known speech or hearing disorder.

Apparatus

White noise was supplied by the same generator described in Experiment I, and fed into subjects' ears through a cushioned headset.

A lavalier microphone was attached to a boom on the headset, thereby controlling for mouth-microphone distance in the recording of voice levels.

A SLM was used to guide subjects' voice levels for the initial autophonic scaling, and its accompanying calibrator was used to record a known calibrating noise on the experimental tape before each session, ready for later measurement of voice levels with a Froekjaer-Jensen level recorder. Vocalisations were recorded on a Revox tape recorder.

A signal light and stopwatch were used to cue responses.

Procedure

(i) Autophonic Scaling

A calibrating noise (97dBA) was first recorded on the experimental tape. The subject was given written instructions as follows, which outline the procedure. (The gaps represent pauses for questions, practice, etc.)

> For your first task, I will ask you to produce 'ah's' at different loudnesses. With each 'ah' I want you to try to centre the needle on the dial of the sound level meter in front of you. Between 'ah's', I shall change the sensitivity of the meter, so that a different loudness will be required each time in order to centre the needle.
>
> Let us try centring the needle a couple of times now, for practice. Begin when the light in front of you flashes. Stop when the light flashes a second time.
>
> ...
>
> Each time you have produced an 'ah', I want you to judge the loudness of it, as follows.
>
> Before and after each 'ah', you will hear a man's recorded voice over the headset. The loudness of his voice, which will always be the same, has the value 10. This is the standard by which you should judge the loudness of your own voice. For instance, if your 'ah' was twice as loud as the standard, write down 20; if it was only half as loud, write down 5. Any type of number may be used - whole number, decimal, or fraction.
>
> I will play the standard to you a few times now. Remember, its value is 10.
>
> ...
>
> Now we will start your attempt to centre the needle on the sound level meter.
>
> First, I will play the standard. Then, when the light flashes in front of you, begin your 'ah'. I shall flash the light again when you have managed to centre the needle for 2 seconds together. At that point, stop vocalizing, listen to the standard again, then write down the loudness value of your voice as you judge
>
> ...

it to have been while the needle was centred. Remember, judge the loudness in relation to the standard, 10.

Centring of the SLM needle was accepted within a 3DB range of the dead centre, held for 2 seconds. Intervals of 4 seconds were left between the subject's hearing the standard, and being cued to begin vocalising; and between the end of the subject's vocalisation and the re-playing of the standard.

Three SLM gain settings were used, 50dBA, 60dBA, and 70dBA, meaning that vocalisations of these magnitudes had to be produced in order to centre the needle. These settings were each used 3 times, in 3 separate randomisations with the restriction that no gain setting followed itself. This procedure resulted in 9 vocalisations from each subject with which to set up the autophonic scale, the mean value of the three vocalisations at each gain setting being used.*

(ii) Loudness Matching and Magnitude Estimation

Subjects were given the following written instructions:

This second task consists of simply imitating the loudness of the standard, but with varying levels of noise coming through the headset. Try to ignore the noise - the important thing for you to concentrate on is matching the loudness of the standard.

It may sometimes happen that you feel you have not matched the standard exactly. In these cases, please write down a number which you feel represents the loudness you did produce. As before, call the standard 10, and decide on the other numbers in relation to the standard. When you are sure you have matched the standard exactly, then write down 10.

...

---

* The last 1.5 second of each vocalisation was actually measured on the Froekjaer-Jensen level recorder.

> I will play the standard to you before and after each match. When you have heard the standard, wait until the signal light flashes, produce your matching loudness, and hold your voice at that loudness until the light flashes again, which will be after about 2 seconds. Then listen to the standard again, and write down your evaluation of the loudness you produced - a perfect match (10), or some other number.

The two noise levels were 65dBA and 75dBA, and the third sidetone condition was simply wearing the headset. There were 6 possible orders of presentation of sidetone conditions, and 2 men and 2 women were assigned at random to each order, within the otherwise repeated measures design. Three loudness matches and magnitude estimates were obtained from each subject for each sidetone condition, and the mean of these three values was used in the analysis of results.*

RESULTS AND DISCUSSION

The first aim of this experiment was to see whether there are any circumstances in which subjects can keep a constant voice level under varying sidetone conditions, and an analysis of subjects' success in loudness matching in different sidetone conditions should give us this information. Accordingly, an analysis of variance was carried out on the voice level measurements obtained in the loudness matching section of the experiment. The result (Table 10) showed a significant effect of sidetone condition on voice level ($F = 34.87$, $df = 2/36$, $p < .001$).

---

* The last 1.5 second of each vocalisation was actually measured on the Froekjaer-Jensen level recorder.

Source	Sum of Squares	df	Mean Square	F	p
Total	3523.45	71			
Sidetone Condition	335.37	2	167.69	34.87	<.001
Order	397.55	5	79.51	<1	
Sidetone x Order	70.34	10	7.03	1.46	
Subjects within Order	2547.08	18	141.50		
Sidetone x Subjects Within Order	173.11	36	4.81		

Table 8: Analysis of Variance on Voice Level Means Experiment V

The meaning of this result is that subjects were not able to match the loudness standard presented to them with the same loudness in each sidetone condition. On the contrary, their voice level varied with the sidetone condition (Fig.6).

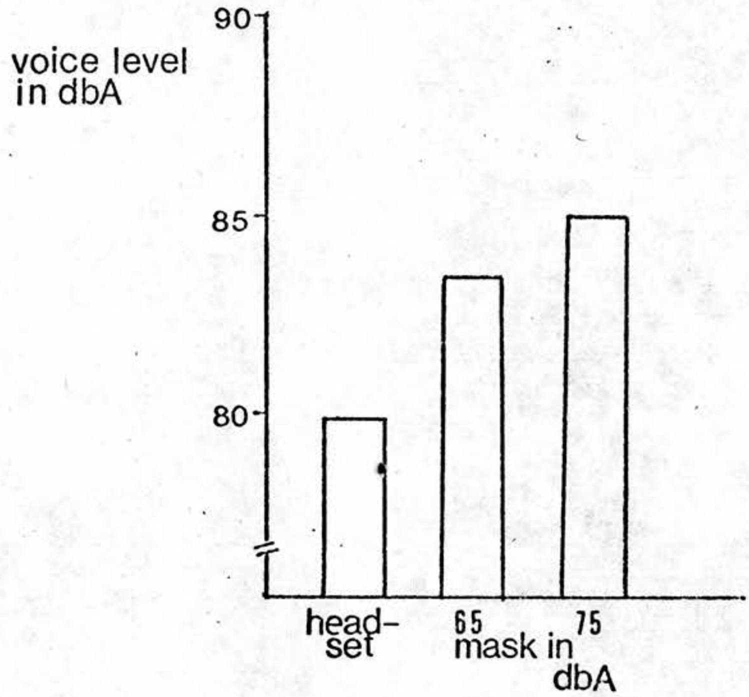

Figure 6: Mean voice levels in different sidetone conditions Experiment V

The second aim of the experiment was to find out whether subjects can judge their absolute voice level under different sidetone conditions, by asking them to judge the accuracy of their attempts to match the loudness standard under the different sidetone conditions. It had been hypothesised that the reduction of sidetone cues by white noise masking would impair subjects' judgment of their own absolute voice loudness. Therefore it was expected that there would be an increasing discrepancy, with noise, between the judgment predicted for any given voice level produced by the subject, and the judgment actually obtained. The predictions of loudness judgments from voice level were obtained from the regression of loudness judgment upon voice level in the initial autophonic scale ($r = .74$, $t = 5.16$, $df = 22$, $p < .001$). An analysis of variance was carried out, then, on the differences between predicted and obtained loudness judgments in the varying sidetone conditions. This showed a significant effect of sidetone condition ($F = 24.04$, $df = 2/36$, $p < .001$; see Table 9).

Source	Sum of Squares	df	Mean Square	F	p
Total	1424.04	71			
Sidetone Condition	100.5	2	50.25	24.04	<.001
Order	152.18	5	30.44	<1	
Sidetone x Order	22.95	10	2.29	1.1	
Subjects Within Order	1073.3	18	59.63		
Sidetone x Subjects Within Order	75.11	36	2.09		

Table 9: Analysis of Variance on Differences between Predicted and Obtained Loudness Judgments of own Voice in Different Sidetone Conditions Experiment V

Sidetone changes, then, significantly affected subjects' judgments of their own voice loudness; but were their judgments affected in the way hypothesised, which was that the discrepancy between predicted and obtained judgment would increase with increasing noise (and therefore decreasing sidetone)? Table 10 and Fig. 7 give this information. It will be seen that, although

Mean Loudness Judgment	Sidetone Condition		
	Headset	65dBA Noise	75dBA Noise
Predicted	10.20	12.42	13.29
Obtained	9.92	10.15	10.24

Table 10: Mean Loudness Judgments Obtained and Predicted (from Initial Autophonic Scale) in Different Sidetone Conditions Experiment V

subjects' judgment of their voice levels was impaired, it was not erratic, but rather systematic, in the manner predicted. That is, there was an increasing discrepancy between objective voice level and judged voice level as sidetone cues were masked by noise; subjects consistently underestimated their absolute voice levels in the absence of normal sidetone cues. There was, however, a slight rise in their loudness estimates as noise level rose, which gives ground for saying that their judgments of relative voice loudness were correct, as Lane et al had previously found.

Clearly, then, subjects were not relying on vocal effort to judge their absolute voice loudness; as their voice levels rose across the sidetone conditions, they continued to judge them as closely matching the loudness of the standard (10). That this is not an artifact of the autophonic scaling range is shown by the fact that whereas the frequency of judgements below 10 was 69, the frequency of judgments above 10 was 110 in the autophonic

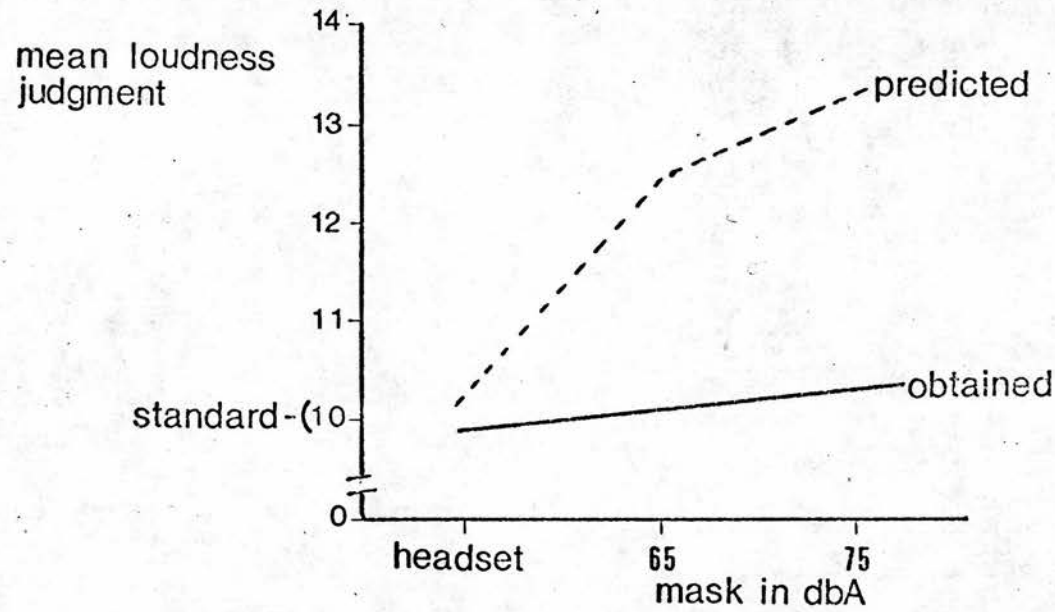

Figure 7: Mean loudness judgments obtained and predicted (from initial autophonic scale) in different sidetone conditions, Experiment V

Loudness
Judgment

2 ''''
3 '''''''''''
4 ''
5 '''''''''''''''''''
6 '''''
7 ''''''
8 ''''''''''''''
9 '''''''''
10 ''''''''''''''''''''''''''''''''''''
11 ''''''''
12 ''''''''''''''
13 ''''
14 ''''
15 '''''''''''''''''
16 '''''''
17 ''''
18 '''''''
19 '
20 '''''''''''''''''''''''''
21
22 ''
23 '
24
25 ''''''''
.
.
.
30 '''''''
.
.
.
40 '

Fig. 8: Frequencies with which Loudness Judgment Numbers were used in Autophonic Scaling (Whole Numbers only - decimals rounded) Experiment V

scaling; thus there was no 'ceiling' at 10. The fact that 10 was also the 'preferred' value in the autophonic scaling (see Fig. 8) suggests that, in the abnormal sidetone conditions, either subjects did not know what voice levels they were producing, and so opted for values clustering around the preferred value most often, or they genuinely believed that they were most often successfully matching the standard. The significant correlation between objective voice levels and judged loudnesses in the autophonic scale disposes of the possibility that subjects merely preferred to assign '10' to any voice level in any part of the experiment.

In conclusion, then, we have some clear evidence in favour of the combined theory. Sidetone is clearly needed for the production, and the judgment, of absolute voice levels; it is <u>not</u> needed for the judgment or production of relative voice levels, which will proceed involuntarily, but systematically, with changes in acoustic conditions.

<u>Interim Summary and Conclusion</u>

Chapter One pointed out that self-hearing was believed to have a role in speech production well before Lombard's demonstration early in this century, and that this belief probably stemmed from two sources: the observation of speech deficits accompanying deafness, and the general notion, disseminated by la Mettrie, that movement is controlled by sensory inflow. It was also pointed out, however, that Lombard's demonstration in itself does not establish that speech is controlled by self-hearing, and neither does the DAF effect demonstrated by Lee in 1950, even though the DAF effect is most often cited in support of the role of self-hearing, just as the Bell-Magendie discovery of separate sensory and motor nerves in the spinal column was taken to establish the role of sensation in movement long after la Mettrie had put this notion into general currency. Both of these ideas

- self-hearing in the control of speech, and self-perception in the control of movement, were given scientific clothing, rather than support, by the later empirical demonstrations. Neither view, however, has gone unchallenged. The Lombard effect has been interpreted as a direct stimulation of the larynx from the ear (cochleo-phonatory reflex), and all sidetone effects have been reinterpreted by Lane and Tranel as expressions of a general strategy to conserve the intelligibility of speech. It has even been argued that physiological provision is made to *attenuate* self-hearing, in order to protect the ear from vibration, or to prevent masking of external sounds.

In Chapter Two, we looked in detail at some of the evidence concerning the role of self-hearing in speech. The motor speech difficulties of the congenitally deaf can be accounted for simply by the fact that they have never had access to normal models of speech, without bringing in the notion of self-hearing. The deterioration of speech following adventitious deafness does, however, indicate that self-hearing is needed for the maintenance of normal speech.

If self-hearing *were* needed to maintain normal speech, what specific features of speech would be under auditory control, and would the auditory control need to be continuous? In Section 3 of Chapter Two, we looked at the general model of sensory inflow control of movement for answers to these questions. It emerged that the features of performance likely to need extrinsic sensory inflow are those which need to be set with reference to anchor-points in the environment - namely, *absolute* rather than relative features of performance. These features needing absolute specification for accurate performance are also *suprasegmental*, such as alignment on the page in writing and drawing, and, in the case of speech, voice level, intonation, and possibly rate.

However, since speech is not a loaded skill, or one which very often meets unpredictable performance conditions, it seemed unlikely that closed-loop control would be a continuous feature of speech. Even so, the performance conditions of speech sometimes do change, for example, with changes in the acoustic environment, such as ambient noise, and therefore it seemed likely that voice level would then come under auditory control.

This whole position then seemed to be undermined by considering Lane and Tranel's theory that the so-called auditory feedback effects on speech do not reflect interference with speakers' self-hearing, but indicate rather that when speakers hear changes in the acoustic environment, they adjust their speech accordingly, in order to remain intelligible to listeners. In other words, speakers then adjust their speech for directly social, rather than merely sensory-motor reasons. According to Lane and Tranel, the adjustment of voice level in these circumstances would be carried out through the monitoring of vocal effort.

It follows from Lane and Tranel's theory that speech affected by auditory manipulations should be more, or at least no less intelligible, than normal speech. The Lombard, Fletcher, sidetone attenuation, and ear-plug effects on voice level were evaluated with this point in mind, and it was clear (i) that none of these effects can be called decrements in performance, in the way that the sensory-motor explanation of them tends to suggest, and (ii) that voice level under these effects changes systematically, and not with random loss of control, as the sensory-motor explanation also implies.

On the other hand, it was pointed out that there are features of speech under these effects which, while they do not rule out a social or intelligibility-preserving function, nevertheless indicate that such a function must be a rather crudely operating one.

It was also pointed out that the Lane-Tranel thesis implies that one's own voice is treated as part of the acoustic environment, and therefore in a sense would be monitored by the speaker, although the occurrence of the Fletcher effect, in which one's own voice when amplified reduces voice level, indicates that one's own voice must be treated as a special percept, distinct from the environment, otherwise it would be treated as interference, and the Lombard effect would be produced. There was some discussion of the difficulty of logically separating the speaker's own voice from the acoustic environment, which led to an attempt to specify all the input-output features in speech. This found that a change in any one input feature (ie environment, air-conducted sidetone, bone-conducted sidetone) would affect at least one other input feature, suggesting that these input features are probably not empirically separable, either.

Nevertheless, the Lane-Tranel theory can be given a substantive interpretation, and the first two experiments in Chapter Three set out to disentangle the two different views of auditory effects on speech - the sensory-motor theory, and the Lane-Tranel social theory, by attempting to bring acoustic and social requirements into conflict for speakers, on the assumption that one requirement would override the other, and thus be shown to be prior. Both experiments instructed speakers that their intelligibility would be optimum if they kept their voice level constant for communication with a listener at a given distance, regardless of what they themselves were hearing (which was interfered with in the first experiment by noise, and in the second by ear-plugs). Speakers responded to these situations with an unexpected and elegant compromise; they graded their voice levels within each distance, in accordance with what they evidently took to be the overall acoustic environment (but which was in fact their private acoustic environment). This result was in keeping with Lane and Tranel's theory, although the response seemed only crudely social when it was considered that (i) speakers ignored the warning, in the first experiment, that if they shouted very loud their intelligibility might be impaired, and (ii) speakers failed to notice that their listeners were not, in fact, in the same acoustic environment as themselves, although they received adequate social feedback about this in the first experiment.

Experiment III, demonstrating the Lombard effect in non-linguistic vocalisation, showed that this effect is not exclusively linked with communication situations.

Although Experiment III had already demonstrated that the Lombard response is not communication-specific, Experiment IV provided some evidence that the response may be communication-linked, since it showed that the masking of the right ear, which preferentially monitors linguistic sounds, raised voice level more effectively than left-ear masking.

Further consideration of the competing sensory-motor and social theories of sidetone effects in the light of the foregoing experiments suggested that a combination of the two theories could overcome the objections to both, and also account for more of the observed facts. This combined theory was that sidetone is needed for the attainment of absolute values in speech, but not for the attainment of relative values. In terms of the Lombard effect, speakers do adjust their voice level as a social response, but the adjustments which can be made on these grounds, without self-hearing, are only relative ones. That is, speakers can change their voice level crudely up or down (perhaps using vocal effort alone) according to the acoustic environment, but do not have fine control over their voice level in the absence of self-hearing.

Experiment V set out to test the predictions, arising from this compromise theory, (i) that subjects would be increasingly less able to match an external standard voice level as their sidetone was increasingly masked by noise, and (ii) that subjects would make increasingly discrepant judgments of their own voice loudness with increasing masking of their sidetone. These predictions were confirmed, leading to the conclusions that self-hearing is needed for fine voice level control, and that whatever social response is contained within the Lombard effect, it is sufficiently crude that while subjects are aware of relative changes in their voice level, they grossly underestimate the absolute voice levels they generate under noise masking. However, in so far as this is a strategy which increases the probability of effective communication, it is an efficient one.

CHAPTER FOUR

SIDETONE EFFECTS ON SPEECH RATE, ARTICULATION AND INTELLIGIBILITY

This far, we have looked at sidetone effects on voice level only. These are clearly direct compensations for the perceived communication conditions (regardless of whether these conditions are perceived by the speaker through self-hearing, or through monitoring of the environment).

Such compensations lead either to the preservation of intelligibility (intelligibility being inferred from increased voice level, as in the case of the Lombard, sidetone attenuation and ear-plugs in quiet effects), or to a reduced voice level more appropriate to the apparent acoustic environment (Fletcher effect).

Of the other sidetone effects we will look at in this chapter, all but one can also be described as compensatory. Two of them, the filtering effect and the larynx sidetone effect, compensate in an indirect fashion which conserves, even enhances intelligibility. Accelerated sidetone does not compensate at all, but instead acts like a positive feedback, since it increases the rate of speech. This sidetone effect produces no apparent change in intelligibility, however. Finally, the sidetone effect which is the best known, delayed auditory feedback, does compensate directly, but in a manner which dramatically impairs intelligibility. Compensation alone, therefore, does not guarantee that intelligibility will be conserved; nor does the absence of compensation necessarily mean that intelligibility will be impaired.

1. The Filtering Effect

Peters (1955) hypothesised that as speakers' sidetone became less intelligible, they would compensate by increasing the intelligibility of their speech. Note that this is a hypothesis which neatly combines a sensory-motor and an intelligibility-conserving view of the function of sidetone: self-hearing is used by speakers to monitor their intelligibility.

Peters varied the intelligibility of his subjects' sidetone by means of high-pass and low-pass filtering, and controlled for the resulting changes in sidetone level. He found that speakers were more intelligible (as judged by listeners in noise) when the sidetone frequencies above 600Hz were attenuated, than they were without filtering.

Lane and van Teslaar (1974) carried out a similar experiment with high- and low-pass filtering of sidetone, again controlling for level effects. They also found that speech was more intelligible (to listeners in noise) when sidetone was filtered. Furthermore, their frequency analyses of the speech produced under this condition showed that there was no direct compensation for the filtering, in the sense that the frequencies attenuated in the sidetone were not pre-emphasised in the speech. However, clearly some intelligibility compensation took place, and Lane and van Teslaar suggest that this was probably at the level of articulation.

2. The Acceleration Effect

Peters (1954) claimed that it was possible to accelerate the return of sidetone to the speaker, though his method of achieving this is not fully described. The result was to increase the rate of speaking, although his subjects rapidly adapted to the change and reverted to their normal rate. Peters did not measure intelligibility. In a similar experiment, Davidson (1959) accelerated sidetone by placing a microphone at a point closer to the lips than the tragus of the ear (ie closer to the lips than 6 inches, and in this case at the corner of the mouth). This procedure also increased the rate of speech, and although in addition it increased the range of intonation, there was no effect on the intelligibility of the speech.

3. A 'Contingency' Explanation of the DAF Effect

The DAF effect is often interpreted as the prime example of a 'compensatory' response to altered sensory feedback. We have seen earlier that this effect is often cited in theories of movement production which stress the role of sensory feedback. Perhaps its appeal for such a purpose lies in the fact that its apparent 'compensatory' nature is so patent: because the feedback is 'missing', the speaker attempts to 'replace' it by repeating or prolonging phonemes, and by speaking louder. Furthermore, the longer the delay, the louder the speech (Black, 1951). The Lombard effect is also usually explained as an attempt to replace the missing feedback by speaking louder, but by the same logic it should also produce repetitions or prolongations of phonemes, which it never does.

As we have already pointed out, the best evidence so far that sidetone has a regulatory, feedback role in speech production is that speech deteriorates after adventitious deafness. However, most writers who attribute this feedback function to sidetone have been convinced rather by the effects on speech of delayed auditory feedback.

As Lane and Tranel have emphasised, DAF effects are almost always the 'clincher' argument for those who write in favour of the control of speech by self-hearing. Thus Davis (1951) after first briefly considering the difficulties of the congenitally deaf in learning to speak, goes on:

> "... 'monitoring' is essential to good normal speech ... we speak much better if we are continually informed as to the success or failure of the nervous messages to our muscles of speech in producing sounds that are actually intelligible to ourselves. The interference produced when the 'sidetone' returning to the talker is artificially delayed, shows dramatically how seriously speech can be disrupted by tampering with this automatic control." (p 5)

Mowrer (1958) writes emphatically about DAF:

> "There could hardly be a more convincing demonstration than this of the fact that even such a highly practiced, over-learned, 'habitual' activity as talking is <u>not</u> a matter of fixated S-R bonds which automatically produce certain responses once the 'switch' is thrown, but instead involves a constant 'monitoring' and control of behaviour AS IT OCCURS, on the basis of the sensory consequences thereof. Obviously and manifestly one LISTENS, very closely, to his own speech." (pp 149-50)

Stanton (1958) refers solely to DAF as the evidence for control of speech by self-hearing:

> "The disturbances of stress and rhythm, the lengthening of vowel sound duration, the faulty articulation and reduplications of consonants, and the retardation of the flow of speech produced by delaying the auditory feedback, all indicate that an important influence in controlling spoken speech must normally be exerted through the auditory pathways." (p 380)

Although she gives some consideration to the effects of adventitious deafness, DAF is clearly the clincher for Harris (1970), too:

> "Anecdotal experience suggests that traumatic deafness, in adulthood, does not cause immediate degradation of speech quality. On the other·hand, continuous acoustic monitoring must have some role in speech maintenance; the speech of deafened adults does apparently deteriorate eventually, although this phenomenon has not been adequately studied. Furthermore, delayed auditory feedback has devastating effects on speech ... In summary, then, continuous auditory monitoring appears to be unnecessary for speech production, but maintenance of normal articulation cannot survive serious distortion or deprivation of auditory feedback." (p 62)

Even those arguing <u>against</u> the notion of auditory feedback control confine their attention almost exclusively to DAF effects:

> "The intervention consists of CHANGING the arrival time of most auditory input, and not eliminating it, and the effects observed may be due to the nature of the changes, rather than to the fact that normal feedback is necessary ... such a time-locked effect would not be predicted simply from the notion that normal feedback is essential to normal production." (MacNeilage and Ladefoged, 1976, p106-7)

Since the argument from DAF holds such a central position, it is worth considering rather closely. Delayed sensory feedback effects are not confined either to speech, or to the auditory modality. Clapping, keytapping (Chase, Harvey, Standfast, Rapin and Sutton, 1959) and playing a musical instrument (Kalmus, Denes and Fry, 1955) are all disrupted by DAF; and other sequential motor tasks can be disrupted by such delayed sensory feedback as a light flash (Cullen and Preston, 1968).

On the other hand, it is not the case that delayed feedback will disrupt <u>any</u> task. Archer and Namikas (1958) found that pursuit rotor performance with a tone signalling 'on target' was not disrupted when the tone was delayed. Altshuler (1967) found negligible effects of DAF of breathing sounds on respiration.

There must, therefore, be some specificity in those delayed sensory feedbacks which do perturb speech and other tasks. One type of specificity could be that the effective feedback is truly continuously integral to performance of the task. If this were so, we would not expect delayed <u>occasional</u> feedbacks, such as Archer and Namikas (above) used, to disrupt a task; nor would we expect an instinctive, 'unskilled' behaviour such as breathing to be dependent on extrinsic feedback.

Another type of specificity in delayed sensory feedback effects could be their CONTINGENCY. Butler and Galloway (1957) found that random delay (ie noncontingent DAF in the form of playback of the speaker's voice from a previous recording of the same material currently being read) did not disrupt fluency. This finding suggests, then, that speech DAF, at least, is disruptive only because it is contingent upon the units of performance. If this is so, then it is important to note that it is not the merely acoustic nature of the sidetone which is perturbing, for in both random and contingent delay the subject hears a signal which is acoustically the same, ie their own voice.*

The next experiment was therefore partly designed to determine whether it is the contingency of any disruptive sidetone which accounts for its disruptiveness.

The second motivation for this experiment was to systematically test the effect of a new type of sidetone. This is one which was found to be disruptive during the course of serendipitous laboratory explorations. I am indebted to Jurek Kirakowski for the idea of returning into the speaker's ears the auditory signal of voice frequency at the larynx which can be obtained from a Laryngograph. When this technique was tried with him as a subject, it had the clearly audible effects of devoicing, denasalisation, monotony, and exaggerated articulation. These effects seemed instantly explicable as sidetone compensation effects, since the audible larynx signal (which sounds like a low hum modulated by intonation) contains information about voicing and intonation, but none about articulation. Thus, the speaker whose normal sidetone is masked by the larynx signal would be expected (on the sensory-motor view) to compensate by attempting to reduce

---

* In fact the sidetone heard during DAF disruption differs from random delay 'sidetone' in so far as it contains the disruptions, but we have to assume that any effect of these is relatively minor. The difficulty of assessing the effect of any distorted feedback per se, independently of its own effects superadded, is however one which deserves consideration.

the relatively increased amount of auditory information from voicing and intonation, while attempting to increase the relatively reduced amount of auditory feedback being received from articulation. Only the denasalisation effect seemed inexplicable in these simple feedback-compensation terms.

The two-fold aim of this experiment, then, was to confirm this new sidetone effect and its phonetic characteristics, and to determine whether its effectiveness depends on its contingency upon the sequential structure of speech, by comparing a contingent form of the sidetone with a noncontingent form.

4. Experiment VI: Contingent vs. Noncontingent Larynx Sidetone

Subjects

There were six men, and six women speakers. All were either postgraduate students, or had received some other form of higher education. They represented a wide variety of English accents, and none had any known speech or hearing disorder.

Apparatus and Materials

A Laryngograph picked up the speaker's voice frequency variation at the larynx (via electrodes) and converted it into an auditory signal. This signal, 'larynx sidetone', was fed into a cushioned headset for binaural return to the speaker.

A constant mouth-to-microphone distance was ensured by the use of a lavalier microphone mounted into a boom attached to the headset.

The passage used for oral reading by the speakers was The Rainbow Passage (Fairbanks, 1960). This passage has been constructed so as to contain all the English speech sounds, in their normal frequencies of occurrence.

Procedure

The first few subjects were asked to experiment with the level of the larynx sidetone while speaking in a moderately quiet voice, and to adjust it so that it just masked the level of their speech as heard from outside the headset. Since this invariably resulted in their setting the sidetone volume at its maximum, it was left at this level for all subsequent subjects.

The next step was to make a recording of the larynx sidetone which would serve later as noncontingent larynx sidetone. This was done while the subject read out a variant of the Rainbow Passage (ie with the order of sentences randomized).

The subject was then randomly assigned to a treatment sequence from a design for three treatments in sequence (Li, 1964, p 216). A repeated measures design was used because individual speakers are the major source of variance in speech intelligibility (Pickett, 1956; House, et al, 1965; Williams and Hecker, 1968). There were six sequences (ie possible orders of three treatments) and each sequence was assigned to both a male and a female subject.

The three treatments were, then, contingent larynx sidetone, noncontingent larynx sidetone, and a control side-tone condition in which the speaker's voice was fed back into the headset without interference. All volume and level settings on the equipment were kept constant for each subject across all three conditions.

RESULTS

Three distinct sets of data were obtained from the speech samples. Judgments on the normality of the speech were obtained (i) from a set ot expert judges (phoneticians) and (ii) from a set of lay judges. Finally, (iii) the intonation range of the speech samples was measured.

(i) Expert Judges

The six expert judges were all postgraduate students of Phonetics and all native English speakers.

They listened to the last half (ie about one minute) of every recorded reading of the Rainbow Passage. The last half of each reading was chosen because it was necessary to sample from the readings in view of their length, and the experimenter's impression was that the audible effects of the contingent larynx sidetone condition, at least, became more evident toward the end of a reading. The three readings from each speaker were presented together, (though in a different random order for each speaker) so that any differences within a speaker across the sidetone conditions could be more easily detected by the judges. Finally, the order of speakers on the judgment tape was randomized, so as to be different from their order of appearance in the experiment and so control for any experimenter practice effect.

On prepared response sheets, the phoneticians were asked first to judge the overall abnormality of each reading, ie to state whether it was 'normal' or 'abnormal'. They were asked then to indicate whether, if there was any overall abnormality, it could be further analysed to lie in any of the following phonetic categories:- articulation, voicing, intonation, nasality. Finally, they were to indicate whether any specific phonetic abnormality consisted of exaggeration, distortion, or reduction of the relevant phonetic feature.. See Fig. 12 for an expert judge's completed response sheet.

The experimental hypothesis expressed in terms of these phonetic categories was that, in the contingent larynx sidetone condition, speech would be judged most abnormal overall, and, in particular, articulation would be exaggerated, while voicing, intonation and nasality would be reduced. No such effects were predicted for non-contingent larynx sidetone, or for the control sidetone condition.

If the abnormality is in one or more of these categories,
write E for exaggerated, R for reduced, or D for distorted, whichever seems most appropriate.

NORMAL	ABNORMAL	ARTICULATION	VOICING	INTONATION	NASALISATION
1 ✓					
2 ✓					
3 ✓					
4 ✓					
5 ✓					
6	✓				E✓
7 ✓					
8 ✓					
9	✓				E✓
10	✓			R✓	
11	✓			R✓	
12	✓			R✓	E✓
13	✓	✓D			E✓
14 ✓					
15	✓				E✓
16 ✓					
17 ✓					
18	✓				E✓
19	✓				E✓
20	✓	✓E, D			
21	✓			✓E	
22 ✓					
23	✓				✓E
24	✓	✓D			✓E
25	✓	✓D		✓E	✓E
26	✓		✓R	✓R	
27 ✓					
28	✓			✓R	
29	✓			✓R	
30	✓	✓E (louder)			
31 ✓					
32 ✓					
33 ✓					
34 ✓					
35	✓	✓stutter D?		✓R	✓E
36 ✓					

Figure 9: Expert judge's response sheet
Experiment VI

Overall Abnormality

When the frequencies of overall abnormality judgments were plotted against the order of occurrence of sidetone conditions, there appeared to be carry-over effects (in speakers). For such cases, Li (1964), the source of the experimental design used here, recommends the use of an analysis of variance for change-over designs which tests for and corrects residual effects (Cochran and Cox, 1957, p 133). This analysis showed that the adjusted residual effects were not significant, while the adjusted direct effects (sidetone conditions) were significant: $F = 10.25$, $df = 2/12$, $p < .01$ (Table 11).

Source	Sum of Squares	df	Mean Squares	F	p
Total*	116.97	35			
Individuals	33.64	11			
Periods within Squares	22.66	8			
Direct effects (unadjusted)	38.89	2			
Residual effects (adjusted)	1.56	2	0.78	0.46	
Residual effects (unadjusted)	5.91	2			
Direct effects (adjusted)	34.53	2	17.27	10.25	<.01
Error	20.22	12	1.685		

* Either 'Direct effects (unadjusted) + Residual effects (adjusted)' or 'Residual effects (unadjusted) + Direct effects (adjusted)', both of which add up to the same total, is excluded from the total sum of squares and total degrees of freedom.

Table 11: Analysis of Variance of Judged Overall Abnormality of Speech as an Effect of Sidetone Condition, Experiment VI

Planned comparisons on the adjusted treatment means (Table 12) showed that the significant difference in sidetone conditions lay between the contingent and the noncontingent conditions ($z = 6.82$, one-tailed $p = .00003$).

Control	Noncontingent	Contingent
2.36	1.28	3.94

Table 12: Adjusted Mean Frequencies of Overall Abnormality Judgments

As was predicted, speech in the contingent larynx sidetone condition was judged most abnormal, but a surprising result was that speech in the noncontingent condition was judged less abnormal than speech in the control condition. Because of this unexpected outcome, a Tukey's test comparing all possible treatment means was carried out to test its significance, but the result was negative.

Apart from assigning simple 'normal' or 'abnormal' judgments to the speech samples, and indicating whether they thought any abnormalities were in the specific phonetic categories of articulation, voicing, etc., the judges had also been asked to indicate whether these specific phonetic abnormalities could be further identified as reduced, distorted, or exaggerated articulation, voicing, etc.

In the event, however, the judges did not make many of these finer judgments regarding reduction, distortion, and exaggeration. Therefore, we report first below, as the major finding, the judgments as to simple abnormality within phonetic categories.

Phonetic Category Abnormality

Agreement among the expert judges in the assignment of overall abnormality judgments to the sidetone conditions is implicitly confirmed in the results of the foregoing analysis of variance.

The judges were also found to agree significantly when the frequencies of their abnormality judgments in respect of phonetic categories were ranked over the sidetone conditions (Kendall's $W = .39, \chi^2 = 25.59$, $df = 11$, $p < .01$).

Articulation was judged more abnormal in the (contingent + noncontingent) conditions combined, than in the control condition (Wilcoxon's test: $N = 10$, $T = 0$, two-tailed $p < .01$; see Table 13). There was no significant difference between the contingent and non-contingent conditions.

Control	Noncontingent	Contingent
3	11	15

Table 13: Total Frequencies of 'Abnormal Articulation' Judgments

In the case of <u>voicing</u>, the contingent condition was again judged most abnormal, and again the noncontingent condition was judged <u>less</u> abnormal than the control condition. By Wilcoxon's test, the contingent condition differs significantly from the noncontingent condition ($N = 8$, $T = 3$, one-tailed $p < .025$), and the control condition differs significantly from the other two conditions combined ($N = 8$, $T = 3.5$, two-tailed $p < .05$). Table 14 shows the total frequencies of 'abnormal voicing' judgments.

Control	Noncontingent	Contingent
13	2	18

Table 14: Total Frequencies of 'Abnormal Voicing' Judgments

For <u>intonation</u>, the same pattern held as for articulation. There was more judgments of abnormality in the combined (contingent + noncontingent) conditions than in the control condition (Wilcoxon's test: $N = 6$, $T = 0$, two-tailed $p = .05$) with no other significant differences (Table 15).

Control	Noncontingent	Contingent
4	9	14

Table 15: Total Frequencies of 'Abnormal Intonation' Judgments

Finally, in the case of nasality, there were no significant differences at all, although the frequency of 'abnormal' judgments fell into the same pattern as for voicing, ie CONTINGENT > CONTROL > NONCONTINGENT (Table 16).

Control	Noncontingent	Contingent
17	11	22

Table 16: Total Frequencies of 'Abnormal Nasality' Judgments

Summing up, then, the expert judges were able to separate speech produced with contingent larynx sidetone from speech produced with noncontingent larynx sidetone, judging speech in the contingent condition to be more abnormal overall, as predicted.

With respect to specific phonetic categories, the contingent condition produced most judgments of abnormal articulation, voicing, intonation, and nasality, again as predicted, although only in the case of voicing was the contingent condition significantly different. Judgments of abnormal articulation and intonation significantly separated the control condition from the combined (contingent + noncontingent) conditions, being fewer in the control condition. The results from the expert judges are summarised in Table 17.

An unexpected feature of these results is that the noncontingent condition was judged <u>less</u> abnormal than the control and contingent conditions in three cases - for overall abnormality, for voicing, and for nasality. Although only one of these differences is significant, this finding has an important bearing on Lane & Tranel's theory of sidetone effects. Discussion of this is deferred until the conclusion of the experiment.

What can be said unequivocally in terms of the experimental hypothesis is that contingent larynx sidetone

Phonetic Categories	Rank Order of Sidetone Conditions in Attracting Abnormality Judgments		
	1	2	3
Overall Abnormality	(contingent)	control	(noncontingent)
Articulation	(contingent	noncontingent)	control
Voicing	([contingent]	(control)	[noncontingent])
Intonation	(contingent	noncontingent)	control
Nasality	contingent	control	noncontingent

Table 17: Summary of Results from Expert Judges, Experiment VI (Conditions separated by brackets are significantly different)

produces speech which is judged overall more abnormal than speech produced with noncontingent larynx sidetone, and that this is in agreement with the expectation that contingent sidetone would be disruptive, whereas non-contingent sidetone would not.

As for the specific 'compensatory' phonetic effects, it had been hypothesised that the contingent condition would produce devoicing, denasalisation, monotony, and exaggerated articulation as direct 'compensations' for the features attenuated or emphasised in the sidetone except for denasalisation, which was predicted only because it had been observed in the pilot subject. The contingent condition did in fact produce most judgments of devoicing, denasalisation,* and monotony (though not of exaggerated articulation), but only one of these differences was significant - that for devoicing, which was significantly different from the noncontingent condition alone ($N = 7$, $T = 0$, one-tailed $p = .01$; see Table 18).

Control	Noncontingent	Contingent
11	0	16

Table 18: Total Frequencies of 'Reduced Voicing' Judgments

(ii) Lay Judges

Lay judges were used in addition to expert judges, because it was felt that if contingent larynx sidetone disrupts speech communication as such, apart from the mere perceptual-motor skill performance of speech (a point which Lane and Tranel would consider highly relevant) then the disruption should be apparent to lay listeners as well as to experts.

The nine lay judges were a heterogeneous group with regard to age and education, but all were native English speakers.

*Although this effect had been predicted solely because it was observed in the pilot subject, we retain some faith in it because Garber (1976) also reports decreased nasal resonance resulting from low-pass filtering of sidetone - virtually the equivalent of our larynx sidetone.

They listened to the same prepared tapes as the expert judges, and were asked to judge whether each speech sample was 'normal', 'slightly abnormal', or 'very abnormal', writing their responses on prepared answer sheets. These judgments were independent, ie the judges were instructed that there was no requirement to use all three response categories in judging the three speech samples from a single speaker (this instruction was found to be necessary in pilot runs). Apart from this, they were given no specific directives other than to discount regional accents as abnormalities, and to rely otherwise on their own judgment.

For analysis, the judgments were assigned scores as follows: 'normal' = 0, 'slightly abnormal' = 1, 'very abnormal' = 2. These scores were summed over speakers and then ranked over the sidetone conditions for each judge (Table 19). A Kendall's coefficient of concordance test on these ranks showed that the agreement between judges was significant ($W = .94$, $s = 120.5$, $K = 8$, $N = 3$, $p < .01$).

Judge	Control	Noncontingent	Contingent
1	3	2	1
2	2.5	2.5	1
3	3	2	1
4	3	2	1
5	3	2	1
6	3	2	1
7	3	2	1
8	3	2	1

Table 19: Lay Judges' Abnormality Judgments, Summed over Speakers and Ranked over Sidetone Conditions.
Rank 1 = Most Abnormal

When the scores were summed over judges and ranked over sidetone conditions for each speaker, the result was less clear-cut (Table 20) but nevertheless the rankings were sufficiently similar to result in a significant Kendall's W ($W = .92$, $s = 66.5$, $\chi^2 = 30.36$, $df = 11$, $p < .01$), showing that the judges' agreement result was not due to agreement on a few individual speakers only.

Speaker	Control	Noncontingent	Contingent
1	1	3	2
2	3	2	1
3	1.5	3	1.5
4	3	1	2
5	3	2	1
6	2.5	1	2.5
7	3	1	2
8	2.5	2.5	1
9	3	2	1
10	3	1	2
11	2	3	1
12	2	3	1

Table 20: Lay Judges' Abnormality Scores, Summed over Judges and Ranked over Sidetone Conditions
Rank 1 = Most Abnormal

When the data were plotted according to the order of occurrence of feedback conditions, there again appeared to be carry-over effects on speakers. However, the previously used Cochran and Cox analysis of variance correcting for residual effects showed a significant effect of sidetone conditions ($F = 5.11$, $df = 2/12$, $p < .05$; Table 21).

Source	Sum of Squares	df	Mean Square	F	p
Total	454.22	35			
Individuals	182.89	11			
Periods within squares	70.66	8			
Direct effects (unadjusted)	96.89	2			
Residual effects (adjusted)	17.35	2	8.67	1.20	
Residual effects (unadjusted)	40.68	2			
Direct effects (adjusted)	73.56	2	36.78	5.11	<.05
Error	86.43	12	7.20		

Table 21 Analysis of Variance on Lay Judges' Abnormality Judgments of Speech in Different Sidetone Conditions, Experiment VI

Planned comparisons on the adjusted treatment means showed that there was a significant difference between the control condition and the combined experimental conditions ($z = 2.26$, one-tailed $p = .01$) and between the contingent and the noncontingent conditions ($z = 2.57$, one-tailed $p = .005$; see Table 22).

Control	Noncontingent	Contingent
11.70	12.53	15.43

Table 22: Adjusted Means of Abnormality Scores assigned to Sidetone Conditions (the higher the mean, the greater the abnormality)

Speech in the contingent condition was judged most abnormal, and speech in the control condition least abnormal, with speech in the noncontingent condition being judged intermediate between the other two conditions. This outcome was in accord with the experimental hypothesis that contingent sidetone interference would be more disruptive to speech than noncontingent sidetone interference.

It is interesting to note that the lay judges significantly separated all three sidetone conditions, whereas the expert judges only made a significant separation between the contingent and noncontingent conditions in terms of overall abnormality. It was to be expected that the two sets of judges would use different criteria, but that the phoneticians would make more, not fewer, discriminations. On the other hand, it is possible that the phoneticians had a stricter criterion for 'abnormality' than the lay judges. While their training is in the direction of ever finer auditory discriminations, they are at the same time discouraged from adopting an invidious judgmental attitude towards different varieties of speech (a standpoint in modern linguistics summed up in the dictum "descriptive, not prescriptive").

(iii) Intonation range

It had been hypothesised that, since contingent larynx sidetone contained relatively exaggerated information concerning the speaker's intonation, one effect of this sidetone manipulation would be to cause speakers to compensate by reducing their intonation variation. In other words, it was expected that speakers would adopt a relative monotone in this sidetone condition. 'Intonation range' is an alternative measure of the 'exaggerated/ reduced intonation' category open to the expert judges. These judges had in fact made most categorisations of 'reduced intonation' in the contingent larynx sidetone condition as predicted, but this result was not significant. Intonation range was measured by means of a write-out facility on the laryngograph, which permitted graphic display of 3-second segments of a speaker's intonation pattern, either directly from electrodes placed on the neck, or from tape playback of a previously-recorded auditory representation of the neck electrode pick-ups. In this analysis, tape-playback was used, neck electrode pick-ups having been recorded along with the speech throughout the experiment.

It was not possible to confine this analysis to the last half of each reading of the Rainbow Passage, as in previous analyses, because there were relatively few three-second segments which were sufficiently noise-free to produce a measurable write-out. Consequently the whole of each reading was used from which to select five random samples for measurement. The range of intonation within each sample was measured in terms of the distance in millimetres between the highest and the lowest point of the intonation write-out, and the final datum was the mean of the five ranges measured within each reading.

Carry-over effects were apparent when these means were plotted against the order of occurrence of the feedback conditions, and so the Cochran and Cox analysis of variance adjusting for residual effects was again used.

There were no significant effects either from the analysis of variance or from planned comparisons on the adjusted treatment means, though the adjusted treatment means were in the predicted direction, ie CONTROL, NONCONTINGENT > CONTINGENT (see Table 23).

Control	Noncontingent	Contingent
13.43	15.06	13.21

Table 23: Adjusted Treatment Means (in millimetres) of Intonation Ranges in Different Sidetone Conditions

Note again that the noncontingent condition is unexpectedly different (although not significantly so) from the control condition.

Conclusion

The experiment has demonstrated a clear disruptive effect of contingent larynx sidetone, as distinct from a noncontingent variant of larynx sidetone, showing that the mere contingency of a disruptive sidetone or 'feedback' may account for its disruptiveness, independently of its acoustic quality. The specific phonetic effects of the contingent sidetone are less clear-cut, and do not adequately bear out the 'compensatory hypothesis'.

An interesting and unexpected feature of these results is the effect of the noncontingent condition. In several cases* this condition produced results which were judged less abnormal even than the control condition (by the expert judges), although only one of these differences (abnormality of voicing) is significant. If this were a real effect, it would be congruent with Lane and Tranel's view that the effect of sidetone manipulations is not a decrement of performance, but to cause the speaker to

* overall abnormality, abnormality of voicing, reduced voicing, distorted voicing, abnormal nasality, and exaggerated nasality.

behave so as to remain intelligible. As we have seen, this view is acceptable in relation to the Lombard effect and filtering effects, though perhaps less so for the Fletcher effect. It does not adequately account for the DAF effect, nor for the contingent larynx sidetone effect demonstrated in this experiment, suggesting that these contingent sidetone effects may still be best explained by the traditional sensory-motor view. It is possible, then, that some types of sidetone effect represent true feedback disruptions, resulting in impaired speech performance, while other types of sidetone have the kind of effect postulated by Lane and Tranel, namely an improvement of speech performance, or intelligibility. In other words, it is possible that the speaker responds to non-contingent, or blanket sidetone manipulations (such as noise, filtering, and non-contingent larynx sidetone) by improving intelligibility, as it were in resistance to a general acoustic interference, but that sidetones which are contingent upon the units of speech production represent an integral interference with speech production, and so impair performance in the way that we see happening with DAF and with contingent larynx sidetone.

This argument depends on our finding, that the non-contingent larynx sidetone produced more abnormal speech than the control condition, being a real one. The next experiment therefore compares the intelligibility of the speech produced in the contingent, noncontingent, and control conditions of this experiment, intelligibility being regarded as a measure more directly relevant to the testing of Lane and Tranel's view.

5. Experiment VII: The Effect of Larynx Sidetone on Speech Intelligibility

INTRODUCTION

The previous experiment suggested the somewhat surprising possibility that the non-contingent larynx 'feedback' condition produced an improvement in the judged quality of speech over the control condition.

Such a finding would be in accordance with a development of Lane and Tranel's theory. As it stands, their theory is simply that interference with a speaker's sidetone causes the speaker to attempt to maintain intelligibility. However, it seems to be a reasonable development of the theory to predict that, in cases where the sidetone interference is not contingent upon the speech skill units, speech intelligibility will actually be improved. DAF is the obvious case where the sidetone is contingent, and an impairment in speech performance is the result. The Lombard effect, however, may be seen as an improvement in speech performance, since voice level is linearly related to intelligibility except at the very highest levels, and it is a striking observation that people with breathy, timid voices will produce an excellent voice projection under the Lombard effect. Similarly, Peters (1955) and Lane and Van Teslaar (1974) found that filtering the speaker's sidetone led to improved intelligibility.*

This experiment, then, set out to compare the intelligibility of the speech produced under different sidetone conditions in the preceding experiment, the experimental hypothesis being that the intelligibility of the speech in the three conditions would fall into this pattern:

NONCONTINGENT > CONTROL > CONTINGENT

* When voice level was controlled for

METHOD

Subjects

There were twenty-seven subjects, who were academics, postgraduate students, or people otherwise educated at the tertiary level. They were divided into 3 equal-sized groups, each group listening to samples of speech from one of the sidetone conditions of the previous experiment. The groups were matched as far as possible with regard to age, sex, linguistic background (some were American), and previous exposure to the stimulus materials. (Within each group, five subjects had had some experience of the Rainbow Passage, either as speakers or as judges in the preceding experiment.) None of the subjects had any known hearing defect.

Apparatus and Materials

The stimuli tested for intelligibility were sentences selected from the speech samples obtained in the larynx sidetone experiment. Twelve sentences were selected at random, from the Rainbow Passage, and each sentence was randomly assigned to a speaker in the larynx sidetone experiment. Each speaker's three utterances of their assigned sentence (ie one utterance in each sidetone condition, since the larynx sidetone experiment used a repeated measures design) were then re-recorded. This procedure controlled for the effect of individual speakers across the different sidetone conditions.

In order to highlight any differences in the intelligibility of the speech across these conditions, it was necessary to make the listening task a difficult one. This was done by setting the playback level of the stimulus sentences rather low (mean level 42dBA, as measured with a sound level meter placed between the earphones of the listener's headset)*and in addition having an ambient noise of 64.5dBA (measured again with the sound level meter placed between the earphones at the listener's location) produced by a ventilation fan

* i.e. with the sound level meter in contact with both headphones

approximating a white noise spectrum. The overall signal-to-noise ratio thus produced was -22.5dBA.

Listeners' response sheets consisted of a set of word-pairs for each stimulus sentence. One word out of each pair actually occurred in the sentence (target). The other word (dummy) had been selected from the remainder of the Rainbow Passage after the stimulus sentences had been removed, (thus controlling for vocabulary probability) and was matched as closely as possible for similarity to the target word, eg.

Target	Dummy
white	high
physicists	physical
ideas	tried
abnormally	apparently

The right-left position of targets and dummies on the response sheets was randomized.

Procedure

The subject heard the twelve stimulus sentences from one sidetone condition over a headset, with the ventilation fan producing ambient noise. After each sentence, the tape-recorder was stopped and the subject marked off the words heard on the response sheet, making a two-way choice between target and dummy words. Subjects who had not taken part in the larynx sidetone experiment were told that the stimulus sentences were taken from a passage called The Rainbow Passage, which was about scientific aspects of rainbows, eg the way the raindrops split white light into colours to form the rainbow, and about the rainbow as it appears in myths and legends.

RESULTS

For each listener, the number of words correctly heard in a sentence were converted into a proportion of the total possible correct for the sentence. The final intelligibility score for each listener was the mean proportion of words correct over the twelve sentences.

Averaging over listeners, the mean intelligibility values for the three sidetone conditions were:

noncontingent	contingent	control
0.79	0.63	0.60

A one-way analysis of variance on the listeners' intelligibility scores showed a significant difference among sidetone conditions ($F = 13.41$, df 2/24, $p < .001$; see Table 24).

Source	Sum of Squares	df	Mean Square	F	p
Between Groups	0.19	2	.095	13.41	< .001
Within Groups	0.17	24	.007		
Total	0.36	26			

Table 24: Analysis of Variance on Intelligibility of Speech in Different Sidetone Conditions Experiment VIII

Since the order of conditions was not as had been hypothesised (ie noncontingent > control > contingent) it was not possible to carry out planned comparisons. Using Dunnett's t test for comparisons between several treatments and a control, however, it was found that the noncontingent condition was significantly more intelligible than the control condition, as hypothesised (Dunnett's $t = 4.75$, 1 control and 2 treatments, 24 df, two-tailed $p < .05$) while the contingent condition did not differ significantly from the control condition.

As was noted earlier, the sound pressure level of the stimulus sentences had been measured in order to determine the overall signal-to-noise ratio of the sentences in ambient noise. From this measurement, it was apparent that the levels of the stimulus sentences were different in the sidetone conditions. Although intelligibility is known to increase systematically with the sound pressure level of speech, an increase in sound pressure level is not necessarily the only source of increased intelligibility. Lane and Van Teslaar (1974) for instance, found that filtering sidetone improved speech intelligibility even when there was no effect on voice level, and they attributed this to more careful articulation. It seemed worthwhile, therefore, to check whether the increased intelligibility in the non-contingent larynx sidetone condition in the present experiment could be explained solely on the basis of increased sound pressure level. This was done by performing an Analysis of Covariance on the sound pressure level (x) and intelligibility (y) data from the twelve stimulus sentences (Table 25). The non-significant F ratio obtained shows that the variance attributable to sound pressure level alone was significantly greater than chance, whereas that attributable to any other factor was not.

Source	df	(yy)	(xy)	(xx)	Regression df	Regression (yy)'	Residual df	Residual (yy)''	Mean Square	F
Treatments	2	0.21	2.31	30.06			2	0.04	0.02	2.98
Error	22	0.18	0.85	17.03	1	0.04	21	0.14	0.007	
Total	24	0.39	3.16	47.09	1	0.21	23	0.18		

Table 25: Analysis of Covariance on Sound Pressure Level (x) and Intelligibility (y) of Stimulus Sentences
Experiment VIII

CONCLUSION

Comparison of the intelligibility of speech produced in the different sidetone conditions shows, then, that speech produced with noncontingent larynx sidetone is more intelligible even than speech produced without any sidetone intervention. This result is in keeping with Lane and Tranel's view that speakers attempt to maintain their intelligibility in responding to sidetone interference, and perhaps more in keeping with a theory that acoustic interference serves not merely to maintain, but to improve the intelligibility of speech.

Although the contingent larynx sidetone resulted in speech which was judged significantly more abnormal than speech produced in the other sidetone conditions in the previous experiment, this experiment found no difference in the intelligibility of speech produced in the contingent larynx sidetone and control conditions. This result, again, is in keeping with the Lane-Tranel view, though only partly so. In spite of contingent larynx sidetone's evident detrimental effect on speech, nevertheless speakers in this condition maintained their intelligibility. Yet the difference between the two types of sidetone interference, contingent and noncontingent, is that one improves intelligibility, while the other merely maintains it. Clearly, not all sidetones have the same type of effect on speech production. However, we can at least put forward the conclusion that all non-contingent, or blanket sidetone effects, enhance intelligibility. Since the presentation of noise during a stutterer's pauses alone suppresses stuttering, and therefore the known therapeutic effect of noise on stuttering is probably not due to masking of the voice (Sutton and Chase, 1961; Webster and Dorman, 1970) it is likely that the usual Lombard effect also is a blanket, or generalised acoustic interference effect, rather than simply a masking effect, and it is in this respect like the non-contingent larynx sidetone effect.

Of course when noise is presented during a speaker's pauses, or non-contingent larynx sidetone is presented, there is more acoustic stimulation overall than with contingent sidetones, in which (apart from DAF) auditory stimulation occurs only during the actual speech. It is possible that the increased sound pressure level of DAF speech is attributable to a superadded 'interference effect' of the delayed sidetone.

While the interpretation of non-contingent sidetone effects is, then, relatively straightforward, the interpretation of contingent sidetone effects remains problematic. It is not possible simply to attribute them to sensory-motor mechanisms. There are two reasons for this. The first is that they have such diverse effects. DAF disrupts performance and intelligibility, even though its effects are of a clearly compensatory nature. Contingent larynx sidetone makes speech sound abnormal, but does not, apparently, impair its intelligibility. Whereas both the Fletcher effect and the sidetone attenuation effect are adaptive. Secondly, there is an ambivalence in current notions of a speech auditory-motor control system which makes it difficult to place certain sidetone effects unequivocally in a 'sensory-motor' category. This ambivalence is exactly equivalent to that in Lombard's view of the Lombard effect. Just as Lombard sometimes interpreted the effect as a loss of control over voice level, but at other times as the effort to better hear oneself by speaking louder, so the sensory-motor view of speech production is sometimes that changing feedback disrupts performance, but at other times that changes in speech performance compensate for changed feedback.

Presumably the reasoning behind a compensatory interpretation is that the compensations are designed to maintain performance. But precisely in the case where compensation is most evident, ie DAF speech, performance is not "maintained", in any normal sense of the word. Sensory-motor explanations, then oscillate between deficit and compensatory interpretations of the phenomena, and so make it difficult to attribute contingent sidetone effects as a whole to the sensory-motor category as a whole. All we can conclude is that contingent sidetones can be disruptive, though not necessarily (cf Fletcher and sidetone attenuation effects) and that,even when they are disruptive - ie make speech sound abnormal, they do not necessarily impair intelligibility,as in the case of contingent larynx sidetone.

Finally, we have to consider the fact that the greater intelligibility of speech in the non-contingent larynx sidetone condition in this experiment was shown to be largely due to increased voice level. This suggests that the intelligibility-enhancing response to blanket interference is a relatively simple one - the plain strategy "speak louder", especially as this is also the strategy behind the Lombard effect. However, since Lane and Van Teslaar (1974) found improved intelligibility with filtered sidetone even when voice level effects were controlled for, it is clear that there must be other contributory factors.

6. Chapter Summary and Conclusion

Most sidetone effects on speech can be seen as 'compensations' for the altered sidetone, either in the sense that the speech effects are replacements or corrections for elements missing or incorrect in the sidetone (eg voice level effects, DAF), or in the sense that the speech simply becomes more generally intelligible, without specific compensations being apparent (eg filtering effect, non-contingent larynx sidetone effect). The first type may be called 'direct', the second 'indirect compensation'.

The occurrence of apparent direct compensation is clearly attractive to sensory-motor theories of movement production, whereas indirect compensation is more in accord with an intelligibility-conserving theory of speech sidetone effects, such as Lane and Tranel's. However, the DAF effect demonstrates that direct compensation alone is no guarantee that intelligibility will be conserved.

Since the DAF effect is the only sidetone effect known to impair intelligibility, and also (apart from its converse, the acceleration effect) the only type of sidetone which is contingent upon speech articulation (the Fletcher effect and the sidetone attenuation effect are also contingent, but on voice level rather than on articulation) it was suggested that it is this articulation - contingent nature of DAF which is responsible for its disruptive power.

This suggestion was generalised by predicting that a contingent form of a new sidetone effect - larynx sidetone - would be more disruptive than a non-contingent form of the sidetone, in Experiment VI. Since this sidetone had apparently produced direct compensatory effects in a pilot subject, these effects were also predicted. The contingency hypothesis was confirmed, but it was not unequivocally demonstrated that the phonetic effects of the contingent larynx sidetone were directly compensatory.

An unexpected incidental finding in Experiment VI was that the non-contingent larynx sidetone appeared to produce less abnormal speech, not only than the contingent sidetone, but than a control condition. In other words, this non-contingent or blanket sidetone appeared to produce indirect compensation in speech to a point where the speech was more intelligible than normal speech.

This finding was tested in Experiment VII by comparing the intelligibility of the speech produced in the different sidetone conditions. It was confirmed that non-contingent larynx sidetone produces speech which is more intelligible than normal speech, and found that this increased intelligibility was largely due to increased sound pressure level of the voice.

In conclusion, it appears that contingent, or time-locked sidetones do no more than maintain, and may actually impair speech intelligibility, while non-contingent, or blanket sidetones enhance intelligibility.

In the course of this thesis, I have pointed out several times that our best evidence that speech is under the control of self-hearing seems to be that motor[1] speech deteriorates following the onset of adventitious deafness (AD). There is, however, an important alternative explanation of this effect, which was foreshadowed earlier (p 25) in a quotation from Dalby (1873). He suggests that the reason why speech deteriorates or ceases in the young child who becomes deaf is that the child forgets what speech is like. The mechanism is the same as that which leads to loss of knowledge of a childhood language when a different language is exclusively used later.[2] In other words, what has been lost is not so much the sound of one's own speech, as the sound of others' speech.

Another possible explanation is that AD speakers reproduce speech as they hear it. In other words, the speech they produce is a faithful reflection of the (distorted) speech they hear (see pp 26-7).

So far, then, we have 3 possible explanations of the nature of AD speech:- (i) loss of the ability to monitor one's own speech through self-hearing (this presupposes that self-hearing is necessary); (ii) loss of familiarity with speech in general (this presupposes that hearing of

---

1 As far as is known, there is no effect of AD on language, although there may be psychiatric or less severe effects. Swinnerton (1956) writes of the novelist Meredith that in old age "... helped by deafness, (he) talked loudly and fantastically in monologue".

2 Little is known about the effect of complete language change within adulthood on knowledge of the first language, although there is some anecdotal evidence. De la Mare (1947) writes of a Missionary to the South Seas who 'went native', and "in four years had practically lost all his English speech". He also mentions the less credible, though intriguing, report that "Marco Polo ... with his father and uncle returned from the splendours of Kublai Kahn to Venice in 1295. It is recorded that when these three famous travellers reached home in their shabby Tartar clothes after four and twenty years' absence they had ... almost lost the use of their native tongue". (p 26)

others is more important than hearing oneself); and finally (iii) loss of the normal sound of speech, but with the retained ability to reproduce speech as it is heard. This position is neutral with regard to self-hearing, but also differs from (ii) in directly correlating the nature of the change in speech with the nature of the change in hearing. In other words, it postulates an ability to produce speech which remains constant, independently of the extent of hearing, but nevertheless treating speech input, whatever its nature, as the model to follow, and so resulting in apparently 'deteriorated' speech. The speech mechanism, and the speech-hearing relationship, remain the same; it is the change in only one term of the 'equation' - speech input - which produces the change in speech performance.

One way of deciding among these explanations is to look at the speech of people who have been socially isolated. These provide us with a natural experiment which separates the variables of self-hearing and hearing of others' speech; since the isolated person hears their own speech while not hearing the speech of others. If, then, the isolated person's speech deteriorates in the same way as the AD person, this will suggest that the AD person's speech deterioration comes about through non-hearing of others' speech, rather than through non-hearing of their own speech.

Of course, there are numerous difficulties in directly comparing AD people and social isolates. First, the claim I have just made that isolated persons hear their own speech may be questioned. In this, they are unlike the deaf, though if it were established that their speech deteriorates in a similar way, this would show that loss of self-hearing was not the reason for it - if, as has just been claimed, social isolates continue to hear themselves. There is, in fact, considerable evidence that they do continue to do so. Although it is almost a cliché of literature that castaways and other solitaries soon resort to thinking aloud and, later, to the creation of imaginary companions with whom to hold conversation, such reconstructions, in addition to having intuitive plausibility, probably derive from genuine reports.

Thus Daniel Defoe's character Robinson Crusoe was based on Alexander Selkirk, a Scottish seaman who was marooned alone on the island of Juan Fernandez, off the cost of Chile, for about four years and four months. According to Richard Steele, who interviewed him after his eventual return to England, "It was his Manner to use stated Hours and Places for Exercises of Devotion, which he performed aloud, in order to keep up the Faculties of Speech" (quoted in Ross, 1965, p 309).

Selkirk was not alone in this fear. De la Mare (op cit) writes that "a friend whose work committed him ... to spending much of his days alone in a very solitary part of the world once told me that in order to be sure of his mind there he used, during his walks abroad, to name aloud the different objects he saw" (pp 167-8).

The scientist and Antarctic explorer Douglas Mawson also reported beginning to think aloud after his last surviving companion had been dead two weeks, and he was struggling to fabricate some kind of snow-shoe which might help him reach his destination. "By this time", he recounts "I was doing a good deal of 'thinking out loud' which, by the way, seemed to give some sort of consolation" (quoted in Bristow and Hinton, 1957, p 312).

Selkirk's explanation for his speaking out loud was to preserve his speech; Mawson's was to help him solve a problem, and to give some comfort. Judith Greene (1975) compares the verbal commentary on their own actions and intentions which is sometimes done by people living alone to the overt egocentric speech in the young child, and Vygotsky (1962) attributes to this egocentric speech a planning function which raises the child's acts to the level of purposive behaviour.

As Mawson hints, another reason why isolates might speak out loud might be to create a sense of human presence for themselves - in other words they attempt to reconstruct the missing social stimulus in the form of the human voice. Evelyn Waugh (1948) writes of the garrulousness of men who have been isolated: "I do not know how the legend originated that the men who administer distant territories are 'strong and silent' ... As for their silence, it seems to vary in exact ratio to their distance from civilization. For silence one must go to the young diners-out of London; men in the wide open spaces are, in my experience, wildly garrulous" (pp 205-6).

A second difficulty in directly comparing the speech of social isolates with that of AD speakers is that the claim that social isolates do not hear others' speech is not strictly true, in the sense that their overt vocalisations are sometimes made in response to auditory hallucinations. There seems to be a tendency for isolates to project human or at least animate identity onto sounds. Paulhan (1866) cites a novel of Turgenev's in which a solitary prisoner 'hears' obscene songs in the sound of squeaking slippers. Similarly, De la Mare (op cit) cites the following from the diary of a Dutch seaman who was marooned for some time between 1725(6) and 1728: "Evil spirits beset him, one of whom becomes embodied and on occasion converses with him. These spirits engage also in a violent quarrel in the dark hours, with a din 'as though there had been a hundred coppersmiths at work with beatings and drummings'" (p 277). That these perceptions were projected onto the sound of the sea is suggested by a later quotation from the diary: "'As I walked upon the Strand, I heard again a very dismal noise of cursing and swearing in my own language'".

Granted that such 'hearing' of 'speech' may occur, however, it still seems likely that continued lack of <u>authentic</u> speech input (apart from one's own) would lead to de-familiarisation with the true sound of speech, which might even affect the perception of the 'projected' speech. Further, it seems likely that AD speakers would experience similar percepts, and would at least be able to evoke auditory images or memories of speech.

In terms of the amount of speech imagined, then, isolates and the adventitiously deaf are probably not very different. But they do differ in the amount of speech they hear - because isolates talk to (or for) themselves, whereas the adventitiously deaf, though they may do this,* would not be able to hear themselves doing it. This, then, is the point at issue: that what matters for the deterioration of speech in AD speakers is not the absence of self-hearing, but the absence of real speech input; and if isolates suffer the same speech deterioration as AD speakers, then the absence of real speech input (independently of one's own) is the cause of this.

Having evaluated the difficulties in comparing AD speech with that of isolates, it remains to present such evidence as we have about the quality of isolates' speech.

According to De la Mare (op cit), Increase Mather "tells of yet another anchoret who was cast away on the isle of Roncador in the Caribbean ... when he was rescued, though only 2 years had gone by, the man having in so long a time conversed only with Heaven, looked very strangely, and was not able at first conference promptly to speak and answer" (p 125).

Lane (1976) mentions three cases of loss of speech following a period of social isolation. Two of these involved complete loss of speech, the first being a girl who at the age of 8 was cut off from a camping party in the Pyrenees by a snowstorm, and who when identified 8 years later was speech-less and remained so for life. The second case was a boy of unspecified age who was kept in a cellar for eighteen months, and who, when he was released "'was shown a cat and a cow and asked what they

---

* I have observed one hard-of-hearing old lady who frequently repeated "Now then, let's see" so softly that she could not possibly have heard it herself. Indeed the very repetition indicated the absence of self-monitoring.

were. He no longer knew ... He does not speak; he no longer knows how to speak'" (p 179). Lane's third case is Alexander Selkirk.

The first two cases are somewhat restricted in generality as they are children; the case of Selkirk is more interesting. Accounts of Selkirk's story, of which there are many, always make much of his speech problem. Dingwall (1951) states that he was "at a loss for words ... Speech came strange to Selkirk, and sustained conversation was a matter of great perplexity" (p 36). Ballard (1967) gives more detail, stating that Selkirk had lost his Fife accent, and missed the ends off words, but after his rescue rapidly recovered his speech. In fact, Captain Woodes Rogers' was the only eye-witness account of Selkirk's condition when he was first rescued, and the relevant part of his account is: "At his first coming on board us, he had so much forgot his Language for want of Use, that we could scarce understand him, for he seem'd to speak his words by halves" (quoted in Ross, 1965, p 306).

The naturalist William Dampier was a member of the crew of the 'Duke' which picked Selkirk up, and we must regard it as a matter of either great regret, or suspicion, that he does not mention Selkirk's speech deterioration in any of his writings; suspicion because it seems that Dampier was the kind of scholar* who would have seen the significance, or at least the interest of Selkirk's difficulty in speaking, if it had been a marked one, and would have at least mentioned it. Dampier was also a member of the crew which picked up William, a Guyanese Indian who had been marooned before Selkirk on the same

---

* "He was in fact a student carrying for the nonce the fuzee and hanger of a buccaneer. In happier days, and with a sounder scientific education, his status in a world cruise might have been that of Darwin on the 'Beagle' (Gray, 1970, p xxxiii; see also Matthews, 1952 and Shipman, 1962).

island* for about the same length of time. According to Kemp and Lloyd (1960) "This man ... was left behind by mistake; he was taken off again by Eaton and Dampier four years later, having nearly lost the gift of speech" (p 48). Kemp and Lloyd's source for this could only have been Dampier himself, since Eaton apparently published nothing, yet Dampier's account (Penzer, 1970) makes no mention of the man's speech or lack of it.

Another reason to be sceptical might be thought to be the possibility that any speech deficit in isolates might be a cognitive, rather than a sensory-motor one. However, the fact that Captain Woodes Rogers made Selkirk his Second Mate minimises the likelihood at the same time of <u>either</u> a cognitive or of a marked speech deficit. It is possible, as Woodes Rogers suggests in his description of Selkirk's speech, that speech deficits in isolates are simply due to lack of practice in speech. Social interaction is in fact a highly skilled activity (Argyle, 1978, Chapter 3) and therefore the events which constitute it take place in rapid succession. Factors such as hardness of hearing, speech impediment, even being too young or too old for the company can easily lead to inability to keep up with the pace, and hence to social exclusion and distress. It is clear, then, that lack of practice alone could easily lead to deterioration of this skill. Such an explanation also accounts for the rapid <u>recovery</u> of speech once practice has been resumed. Apparently no-one but Alex Hamilton (1979) noticed Sir Francis Chicester's speech deficit after his solo voyage: "Ever since I met the late Sir Francis Chicester coming off the boat after his epic single-handed journey and found that he'd more or less lost the capacity to speak in consecutive phrases, I've had doubts about going to sea under sail." (p 10)

---

* Juan Fernandez appears to have had more than its fair share of solitary denizens. Dampier (Penzer, op cit) also mentions the case of a sole survivor of a shipwreck on the island who had been isolated there for five years, before William's stay, but again says nothing about his speech.

Therefore recovery in this case must have been very rapid.

Many of the studies on sensory deprivation mention speech difficulties in the subjects (Freedman and Greenblatt, 1960; Mendelson et al, 1961; Myers et al, 1962; Zubek and MacNeill, 1967) but again it is not clear from the reports whether cognitive or motor speech difficulties are meant. Only one description is specific enough to mention "slurred" speech (Cohen et al, 1961).

Although the evidence is not as detailed as could be desired, there seems to be sufficient corroboration among these reports to indicate that there is <u>some</u> speech deficit in isolates, and that the deficit is not entirely a cognitive one, but has purely motor elements - Selkirk's truncated words and the sensorily deprived subjects' slurred speech. The evidence is enough, however, to prevent us from accepting without question that the quality of AD speech is due entirely to the absence of self-hearing, and this conclusion undermines what appeared to be the best evidence available that speech is regulated by self-hearing.

CHAPTER SIX

CONCLUSION

The question posed by this thesis was: what is the role of self-hearing in speech production? and the question was stimulated by Lane and Tranel's novel claim that self-hearing has no role.

The answer to this question is that self-hearing does have at least one important function, which is to enable speakers to compare their voice level with ambient noise level, and thus to set an appropriate voice level.

The work reported has been mainly concerned with the control of voice level, but it has also emerged, in the course of the argument and from the last two experiments, that there is an important difference between the effect on <u>articulation</u> of contingent (or time-locked) sidetone manipulations, and the effect of non-contingent manipulations. Whereas contingent sidetone manipulations disrupt articulation, non-contingent ones actually improve intelligibility. However, this improvement in intelligibility is due to <u>adding</u> interference to the speech task, rather than to simply removing or changing auditory feedback. It is this added interference which causes speakers to become more intelligible, by forcing them to attempt to overcome it.

There are two important distinctions, then, to be borne in mind. First, that between genuine auditory feedback effects, and effects caused by the imposition of auditory interference. Secondly, that between voice level control and articulation control. This second distinction is important because voice level and articulation, it is argued here, are normally under different modes of auditory control.

Regarding voice level control, the work reported here indicates that one function of self-hearing is to set appropriate voice level by comparing voice level with ambient noise level. This comparison activity must be almost continuous during speech, since ambient acoustic conditions are frequently fluctuating, and therefore there must be a closed-loop, audio-motor control system for the regulation of voice level.

At the same time, however, it is argued that the sensory-motor control model of speech production, stimulated by the discovery of the DAF effect, is false. Furthermore, it is argued that voice level control can, in some conditions, be automatic (open-loop) rather than closed-loop.

These conclusions are not as paradoxical as they may at first appear. The role of self-hearing in setting voice level against ambient noise level is clearly supported by the results of Experiment V. This experiment showed that as the level of masking noise increased, and hence as the level of available sidetone decreased, subjects became less able to match an objective standard voice level, and instead raised their voice level in step with the noise. Their subjective judgments of their own voice level, although rising slightly, did not accurately reflect this involuntary increase, but progressively and grossly underestimated it with increasing masking. Therefore vocal effort alone, which Lane and Tranel claim is the cue to autophonic level, is clearly not sufficient, and self-hearing must be the major cue.

Voice level, then, is under closed-loop control, and since ambient noise levels fluctuate frequently during speech, this control must be virtually continuous.

Yet the simple sensory-motor, closed-loop control model of speech which was derived from studies of the DAF effect must be rejected. It must be rejected for several reasons, but one of them we will discuss now is precisely that voice level can, in some conditions, be regulated <u>without</u> self-hearing. I refer here to the Lombard effect, the involuntary raisigg of voice level in step with masking noise. The involuntary nature of this effect is clearly seen in Experiment V, where, although subjects had been instructed to match an external standard, they nevertheless raised their voice level in step with the masking noise. In this Experiment, the Lombard effect was dysfunctional, but in other circumstances the Lombard effect is an appropriate one, as indeed Lane and Tranel argue.

The Lombard effect was previously interpreted as evidence for the closed-loop control of voice level, since the raising of the voice was seen as a loss of control over voice level, due to the masking of auditory feedback. This interpretation fails, however, to take account of the systematic way in which voice level rises with noise level. 'Loss of control' is not an adequate description of this systematic response. The Lombard effect rather reflects an open-loop or automatic mode of vice level control - clearly it cannot be closed-loop,since the speaker's auditory feedback is masked.

There are reasons, firstly for regarding the Lombard effect as functional, and secondly for regarding it as the reflection of an automatic mode of control. Firstly, to speak louder in a noisy environment is an adaptive response, as is the raising of voice level in step with noise level. Other evidence supporting this functional view of the Lombard effect has been cited earlier inthe thesis, but the major further piece of evidence is that the effect is stronger when a speaker is actually communicating, rather than reading out word lists.

Secondly, I have shown in Experiment I that speakers can make adjustments to their voice level even <u>within</u> the Lombard effect. In this experiment, speakers were able to match voice level to distance even though they were already under the Lombard effect, and had no access to auditory feedback. Furthermore, the systematic relations among voice level, distance, and noise level were retained as the noise level increased. This was the case even though subjects had disobeyed the instructions, which were to keep voice-to-distance relations the same as in a no-noise condition. This experiment shows, then, that while speakers cannot maintain absolute voice levels without self-hearing, they can, nevertheless, adjust voice level in a quasi-functional manner, and since these speakers had no auditory feedback available, they must have been operating an open-loop mode of control.

The open-loop nature of this control is the key to its only <u>quasi</u>-functional outcome. In Experiment I, the compensation in voice level for noise level was only of the order of 4dB in voice level for 25dB of noise. Here the speakers were transmitting word lists to listeners, but even in live communication conditions the compensation has never been found to be greater than half way (Lane and Tranel, 1971). However, in spite of its poor absolute compensation for noise, the Lombard effect must be regarded as functional in so far as it raises the <u>probability</u> of successful communication in live conditions. The matching of life-forms to environmental conditions does not proceed in other than a probabilistic (and infinitely slower) fashion, yet evolution is regarded as a functional process. Similarly, the hedgehog's rolling into a prickly ball or the cow's swiping of its tail in response to flies are crude and open-loop in respect of the actual or potential attacker, and are somewhat ineffectual responses; nevertheless they go some way toward achieving the desired result.

Open-loop regulation of voice level, as demonstrated in the first experiment, is one reason for rejecting the closed-loop model of speech production derived from DAF studies. There is another important reason, which is that the DAF effect is an anomaly among side-tone effects. It is the only one which clearly impairs speech performance. The reason for this, as was argued in Experiment VI, seems to be the specifically contingent or time-locked relation of the feedback to articulatory movements. It interferes with auditory feedback in a manner over and above the simple removal of it, and therefore does not, of itself, demonstrate that the presence of auditory feedback is necessary for speech. Indeed, the Lombard effect, which clearly demonstrates that auditory feedback is dispensable for articulatory movements, had already been known for almost forty years before the DAF effect was discovered.

The contingent nature of the feedback also explains the effect of contingent larynx side-tone (reported in Experiments VI and VII) in producing speech which is judged to be abnormal. This was shown by contrasting its effect with that of non-contingent larynx side-tone, which is acoustically identical to it and differs only in its temporal relation to articulatory movements, yet produces highly intelligible speech.

Different 'auditory feedbacks', then, divide into two clear types - those which impair performance, and those which actually improve it. Those which improve performance are not so much 'feedbacks' as interferences, since what they demonstrate is not the use of auditory feedback but rather the ability of speakers to adapt to interference even in the absence of normal access to auditory feedback. Lane and Tranel seem to have based their whole argument against the role of self-hearing on this latter type of effect. By ignoring the necessity for speakers to continuously monitor their voice level for optimum control, they have erred in the reverse direction to the 1950's sensory-motor control theorists, whose error was to conclude that articulation needed to be under continuous auditory monitoring merely because of the anomalous effects of DAF.

Lane and Tranel's contribution, however, has been to point out the much-neglected communicative, and not simply motor-skill, nature of speech. It was as if the early theorists' preoccupation with sensory-motor control, and what they saw as removal-or-manipulation experiments, blinded (or deafened) them to the possibility, and even to the presence, of enhanced speech performance.

Finally, it remains to ask whether the model of auditory-vocal monitoring put forward in this thesis can deal with the effects of adventitious deafness on speech - effects which seemed to be incontrovertible evidence of the need for self-hearing in speech production. To briefly recap this model, it includes continuous closed-loop auditory monitoring of voice level, and open-loop auditory control of articulation, though with intermittent checking of the latter to prevent drift. Deaf speakers' difficulty in setting an appropriate voice level is well accounted for by the implication from this model that they lack the continuous closed-loop control system which is available to the hearing. The fact that articulation deteriorates only gradually after the onset of deafness is also explained very well by the postulated intermittent auditory check on articulation possessed by the hearing. However, as was pointed out in Chapter Five on speech and social isolation, it is very likely that prolonged defamiliarisation with the sound of speech would also be a very strong factor in the slow deterioration of speech, by gradually eroding knowledge of articulatory targets.

Penn's (1955) findings that the nature of the speech deficit is different in perceptive and conductive deafness is compatible with this model. He found that there were more deviations from articulatory targets in perceptive deafness, and that speech in conductive deafness was devoiced and denasalised. In conductive deafness, it is only the conductive mechanism in the middle ear which is damaged, meaning that, although air-borne sound can no longer be heard, bone-conducted vibrations can still be transmitted directly to the cochlea. Conductively deaf speakers can therefore still hear some of their own vocalization via bone conduction. Bone-conducted sound is low in frequency and the fact that their speech is devoiced and denasalised suggests that a compensation is occurring similar to that observed in Experiment VI, in which contingent larynx sidetone (subjectively, a low hum modulated by intonation) produced speech which phoneticians judged to be devoiced and, (though to a less significant degree) denasalised.

Compensations of this type (which include the DAF effect) are due to contingent alterations of sidetone which are anomalous in the sense that, unlike non-contingent alterations, they are not naturally-occurring, and, they impair intelligibility. Although contingent sidetone manipulations can be used to 'drive' speech quite readily, they do not throw much light on the natural communicative process, which, as we have seen, can proceed effectively without self-hearing at least in the short term, and is adaptable to noncontingent acoustic interference in such a way that intelligibility is protected. This adaptability is more relevant to the natural process than the effects of contingent manipulations on speech.

There is however the intriguing possibility that the contingent effects would actually be more intelligible to a listener in the 'same' acoustic environment as the speaker, if such a thing could be arranged, but this is a matter for further research.

BIBLIOGRAPHY

Abbs, J.H. & Eilenberg, G.R. (1976) Peripheral mechanisms of speech motor control. In Lass, N.J. (Ed.) CONTEMPORARY ISSUES IN EXPERIMENTAL PHONETICS, Academic Press, New York.

Abbs, J.H. & Smith, K.U. (1970) Laterality differences in the auditory feedback control of speech. JOURNAL OF SPEECH AND HEARING RESEARCH, 13, 298-303.

Allen, G.D. (1971) Acoustic level and vocal effort as cues for the loudness of speech. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 49, 1831.

Altschuler, M.W. (1967) Negligible effects on respiration of delayed auditory feedback of breathing noise. JOURNAL OF AUDITORY RESEARCH, 7, 181-185.

Annett, J., Morris, P., Holloway, C., & Roth, I. (1974) HUMAN INFORMATION PROCESSING, PART 1, Open University Press, Milton Keynes.

Arbib, M.J. (1972) THE METAPHORICAL BRAIN, Wiley, New York.

Archer, E.J. & Namikas, G.A. (1958) Pursuit rotor performance as a function of delay of information feedback. JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 56, 325-327.

Argyle, M. (1978) THE PSYCHOLOGY OF INTERPERSONAL BEHAVIOUR, Penguin, Harmondsworth.

Arnott, N. (1827) ELEMENTS OF PHYSICS OR NATURAL PHILOSOPHY, GENERAL AND MEDICAL, Underwood, London.

Atkinson, C.J. (1952) Vocal responses during controlled aural stimulation. JOURNAL OF SPEECH AND HEARING DISORDERS, 17, 419-426.

Atkinson, C.J. (1953) Rate and sound pressure level of speaking as affected by spectrum changes in amplified sidetone. STATE UNIVERSITY OF IOWA, SIGNAL CORPS CONTRACT No. DA 36-309 SC-42562, SUPPLEMENTARY REPORT No. 1.

Bain, A. (1859) THE EMOTIONS AND THE WILL, Longmans Green, London.

Baken, R.J. (1969) Distribution of neuromuscular spindles in intrinsic muscles of a human larynx. Ph.D. dissertation, Columbia University.

Bakker, D.J. (1967) Left-right differences in auditory perception of verbal and non-verbal material by children. QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 19, 334-336.

Bakker, D.J. (1968) Ear asymmetry with monaural stimulation. PSYCHONOMIC SCIENCE, 12, 62.

Ballard, M. (1967) THE MONARCH OF JUAN FERNANDEZ, Constable, London.

Barany, E. (1938) A contribution to the physiology of bone conduction. ACTA OTOLARYNGOLOGICA, SUPPLEMENT 26.

Barlow, N (Ed.) (1958) THE AUTOBIOGRAPHY OF CHARLES DARWIN, Collins, London.

von Bekesy, G. (1949) The structure of the middle ear and the hearing of one's own voice by bone conduction. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 21, 217-232.

von Bekesy, G. (1962) The gap between the hearing of external and internal sounds. In Society for Experimental Biology, Symposium No. 16, BIOLOGICAL RECEPTOR MECHANISMS, Cambridge University Press.

Bell, C. (1830) THE NERVOUS SYSTEM OF THE HUMAN BODY, Longmans and J. Taylor, London.

Beloff, J. (1973) PSYCHOLOGICAL SCIENCES, Crosby, Lockwood, Staples, London.

Bernstein, N.A. (1967) THE COORDINATION AND REGULATION OF MOVEMENTS, Pergamon, Oxford.

Birch, H.G. (1956) Experimental investigations in expressive aphasia. NEW YORK STATE JOURNAL OF MEDICINE, 56, 3849.

Birch, H.G. & Lee, J. (1955) Cortical inhibition in expressive aphasia. ARCHIVES OF NEUROLOGY AND PSYCHIATRY, 74, 514.

Black, J.W. (1950a) The effect of room characteristics upon vocal intensity and rate. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 22, 174-176.

Black, J.W. (1950b) Some effects upon voice of hearing tones of varying intensity and frequency while reading. SPEECH MONOGRAPHS, 17, 95-98.

Black, J.W. (195oc) Some effects of auditory stimuli upon voice. JOURNAL OF AVIATION MEDICINE, 21, 251-259.

Black, J.W. (1951) The effect of delayed side-tone upon vocal rate and intensity. JOURNAL OF SPEECH AND HEARING DISORDERS, 16, 56-60.

Black, J.W. (1954) Systematic research in experimental phonetics: 2 signal reception: intelligibility and side-tone.. JOURNAL OF SPEECH AND HEARING DISORDERS, 19, 140-146.

Black, J.W. & Tolhurst, G.C. (1956) Intelligibility as related to the path of airborne side-tone. JOURNAL OF SPEECH AND HEARING DISORDERS, 21, 173-178.

Boone, D.R., (1959) Communication skills and intelligence in right and left hemiplegics. J. SPEECH AND HEARING DISORDERS, 24, 241-248.

Borg, E. & Zakrisson, J.E. (1975) The activity of the stapedius muscle in man during vocalization. ACTA OTOLARYNGOLOGICA, 79, 325.

Boring, E.G. (1942) SENSATION AND PERCEPTION IN THE HISTORY OF EXPERIMENTAL PSYCHOLOGY, Appleton-Century-Crofts, New York.

Bosatra, A., Russolo, M., & Semerano, A. (1975) Tympanic muscle reflex elicited by electric stimulation of the tongue in normal and pathological subjects. ACTA OTOLARYNGOLOGICA, 79, 334-338.

Brett, G.S. (1912) HISTORY OF PSYCHOLOGY, VOL 11, G. Allen, London.

Bristow, J., & Hinton, A. (Eds.) (1957) GREAT TRUE ADVENTURES IN ICE AND SNOW, Arco, London.

Broadbent, D.E. (1952) Speaking and listening simultaneously. JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 43, 267-273.

Browne, L., & Behnke, E. (1884) VOICE, SONG, AND SPEECH, Sampson Low, Marston, Seale and Rivington, London.

Butler, R.A., & Galloway, F.T. (1957) Factorial analysis of the delayed speech feedback phenomenon. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 29, 632-635.

Campbell, B. (1971) The roots of language. In Morton, J. (Ed.) BIOLOGICAL AND SOCIAL FACTORS IN PSYCHOLINGUISTICS, Logos, London.

Cannon, W.B. (1932) THE WISDOM OF THE BODY, Norton, New York.

Carhart, R. (1947) Conservation of speech. In Davis, H. (Ed.) HEARING AND DEAFNESS, Murray Hill, New York.

Carmel, P.W. & Starr, A. (1963) Acoustic and non-acoustic factors modifying middle-ear muscle activity in waking cats. JOURNAL OF NEUROPHYSIOLOGY, 26, 598.

Cassells, J.P. (1883) POLITZER'S TEXT-BOOK OF THE DISEASES OF THE EAR, Baillere, Tindall, and Cox, London.

Chang-Yit, R., Pick, H.L., & Siegel, G.M. (1975) Reliability of sidetone amplification effect in vocal intensity. JOURNAL OF COMMUNICATION DISORDERS, 8, 317-324.

Chase, R.A., Harvey, S., Standfast, S., Rapin, I., & Sutton, S. (1959) Comparison of the effects of delayed auditory feedback on speech and key tapping. SCIENCE 129, 903-904.

Chase, R.A., Sutton, S., & Rapin, I. (1961) Sensory feedback influences on motor performance. JOURNAL OF AUDITORY RESEARCH, 1, 212-222.

Chase, R.A., Cullen, J.K., Niedermeyer, E.F.L., Stark, R.E., & Blumer, D.P. (1967) Ictal speech automatisms and swearing: studies in the auditory feedback control of speech. JOURNAL OF NERVOUS AND MENTAL DISEASES, 144, 406-420.

Cherry, E.C., Sayers, B. McA., & Marland, P.M. (1955) Experiments on the complete suppression of stammering. NATURE, 176, 874.

Cherry, E.C., & Sayers, B. McA. (1956) Experiments upon the total inhibition of stammering by external control and some clinical results. JOURNAL OF PSYCHOSOMATIC RESEARCH, 1, 223.

Cherry, E.C. (1957) ON HUMAN COMMUNICATION, Wiley, London.

Chomsky, N. (1968) LANGUAGE AND MIND, Harcourt, Brace, and Jovanovich, New York.

Cohen, S.I., Silverman, A.J., Bressler, B., & Shmavonian, B. (1961) Problems in isolation studies. In Solomon, P., et al (Eds.) SENSORY DEPRIVATION, Harvard University Press, Cambridge, Massachusetts.

Connolly, K. (1977) The nature of motor skill development. JOURNAL OF HUMAN MOVEMENT STUDIES, 3, 128-143.

Craik, K.J.W. (1947) Theory of the human operator in control systems l. The operator as an engineering system. BRITISH JOURNAL OF PSYCHOLOGY, 38, 56-61.

Craik, K.J.W. (1948) Theory of the human operator in control systems 11. Man as an element in a control system. BRITISH JOURNAL OF PSYCHOLOGY, 38, 142-148.

Cullen, J.K., & Preston, M.S. (1968) Delayed feedback of an amplitude modulated tone. ANNUAL REPORT, NEUROCOMMUNICATIONS LABORATORY, DEPARTMENT OF PSYCHIATRY AND BEHAVIORAL SCIENCES, THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE, 317-330.

Curry, F.K.W., (1967) A comparison of right handed and left handed subjects on verbal and non-verbal dichotic listening tasks. CORTEX, 3, 343-351.

Dalby, W.B., (1873) DISEASES OF THE EAR, Churchill, London.

Daniloff, R.G. (1973) Normal articulation processes. In Minifie, F., Hixon, T.J., & Williams, F. (eds.) NORMAL ASPECTS OF SPEECH, HEARING AND LANGUAGE, Prentice-Hall, Englewood Cliffs, New Jersey.

Davidson, G.D. (1959) Sidetone delay and reading rate, articulation, and pitch. JOURNAL OF SPEECH AND HEARING RESEARCH 2, 266-270.

Davis, H. (1951) Auditory communication. JOURNAL OF SPEECH AND HEARING DISORDERS, 16, 3-8.

De la Mare, W. (1947) DESERT ISLANDS, Faber, London.

Denes, P.B., & Pinson, E.N. (1963) THE SPEECH CHAIN, Bell Telephone Company, New Jersey.

Deutsch, J.A., & Clarkson, J.K. (1959) Nature of the vibrato and the control loop in singing. NATURE, 183, 167-8.

Dingwall, G. (1951) THE STORY OF ALEXANDER SELKIRK, THE TRUE 'ROBINSON CRUSOE', Fifeshire Advertiser, Kirkcaldy.

Dolch, J.P., & Schubert, E.D. (1955) Your own voice under headphones. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 27, 1001.

Elliott, L.L., & Niemoller, A.F. (1970) The role of hearing in controlling voice fundamental frequency. INTERNATIONAL AUDIOLOGY, 9, 47-50.

Espir, M.L.E., & Rose, F.C. (1970) THE BASIC NEUROLOGY OF SPEECH, Blackwell, Oxford and Edinburgh.

Fairbanks, G. (1954) Systematic research in experimental phonetics I: A theory of the speech mechanism as a servosystem. JOURNAL OF SPEECH DISORDERS, 19, 133-139.

Fairbanks, G. & Guttman, N. (1958) Effects of delayed auditory feedback upon articulation. JOURNAL OF SPEECH AND HEARING RESEARCH, 1, 1-11.

Fairbanks, G. (1960) VOICE AND ARTICULATION DRILLBOOK, (Second Edition) Harper and Row, New York.

Flanagan, J.L. (1951) Effect of delay distortion upon the intelligibility of speech. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 23, 303.

Fletcher, H. (1929) SPEECH AND HEARING, Van Nostrand, London.

Fletcher, H., Raff, G.M., & Parmley, F. (1918) Study of the effects of different amounts of sidetone in the telephone set. WESTERN ELECTRIC COMPANY REPORT NO. 19412, CASE NO. 120622.

Fournié, E. (1866) PHYSIOLOGIE DE LA VOIX, Delahaye, Paris.

Freedman, S.J., & Greenblatt, M. (1960) Studies in human isolation II. Hallucinations and other cognitive findings. U.S. ARMED FORCES MEDICAL JOURNAL, 11, 1479-1497.

Fry, D.B. (1954) The experimental study of speech. NATURE, 173, 844-846.

Fry, D.B. (1957) Speech and language. JOURNAL OF LARYNGOLOGY AND OTOLOGY, 7, 434-452.

Fry, D.B. (1966) The development of the phonological system in the normal and deaf child. In Smith, F., & Miller, G.A. (Eds.) THE GENESIS OF LANGUAGE, M.I.T. Press, Cambridge, Massachusetts.

Garber, S. (1976) The effects of feedback filtering on nasality. AMERICAN SPEECH AND HEARING ASSOCIATION CONVENTION, Houston, Texas.

Garber, S., Siegel, G.M., Pick, H, & Alcorn, S.R. (1976) The influence of selected masking noises on Lombard and sidetone amplification effects. JOURNAL OF SPEECH AND HEARING RESEARCH, 19, 523-535.

Garde, E.J. (1965) Le réflexe cochléophonatoire: Etat actuel de la question d'après quelques faits expérimentaux et pathologiques. REVUE LARYNGOLOGIQUE DE BORDEAUX, 86, 631-638.

Gaynor, E.B. (1974) Humming and nonorganic hearing loss. ARCHIVES OF OTOLARYNGOLOGY, 100, 199-200.

Gibson, J.J. (1966) THE SENSES CONSIDERED AS PERCEPTUAL SYSTEMS, George Allen and Unwin, London.

Glencross, D.J. (1977) Control of skilled movements. PSYCHOLOGICAL BULLETIN, 84, 14-29.

Goldman-Eisler, F. (1954) On the variability of the speed of talking and on its relation to the length of utterances in conversation. BRITISH JOURNAL OF PSYCHOLOGY, 45, 94-107.

Goldman-Eisler, F. (1968) PSYCHOLINGUISTICS, Academic Press, London.

Gray, Sir A. (1970) Introduction to Penzer, N.M. (Ed.) A NEW VOYAGE ROUND THE WORLD. BY WILLIAM DAMPIER, Da Capo Press, New York.

Greene, J. (1975) THINKING AND LANGUAGE, Methuen, London.

Greenwald, A.G. (1970) Sensory feedback mechanisms in performance control. PSYCHOLOGICAL REVIEW, 77, 73-99.

Hamilton, A. (1979) On the rack. GUARDIAN, 11 August, 1979.

Hardcastle, W.J. (1975) Some aspects of speech production under controlled conditions of oral anaesthesia and auditory masking. JOURNAL OF PHONETICS, 3, 197-214.

Harris, K.S. (1970) Physiological aspects of articulatory behavior. HASKINS LABORATORIES STATUS REPORTS ON SPEECH RESEARCH, 23, 49-67.

Hebb, D.O., Heath, E.S., & Stuart, E.A. (1954) Experimental deafness. CANADIAN JOURNAL OF PSYCHOLOGY, 8, 152-156.

Held, R. (1965) Plasticity in sensory-motor systems. SCIENTIFIC AMERICAN, 213, 89-94.

Holmes, C., & Holtzman, P. (1966) Effect of white noise on disinhibition of verbal expression. PERCEPTUAL AND MOTOR SKILLS, 23, 1039-1042.

House, A.S., Williams, C.E., Hecker, M.H.L., & Kryter, K.D. (1965) Articulation testing methods: consonantal differentiation with a closed response set. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 37, 158-166.

Husson, R. (1962) PHYSIOLOGIE DE LA PHONATION, Masson, Paris.

Jen, P. H.-S., & Suga, N. Coordinated activities of middle-ear and laryngeal muscles in echolocating bats. SCIENCE 191, 950-952.

Jepsen, O. (1963) Middle-ear muscle reflexes in man. In Jerger, J. (Ed.) MODERN DEVELOPMENTS IN AUDIOLOGY, Academic Press, New York.

Kalmus, H., Denes, P., & Fry, D.B. (1955) Effect of delayed acoustic feed-back on some non-vocal activities. NATURE, 175, 1078.

Karlovich, R.S., & Luterman, B.F. (1969) Auditory fatigue during articulation: preliminary investigations. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 45, 786-787.

Kemp, P.K., & Lloyd, C. (1960) BRETHREN OF THE COAST, Heinemann, London.

Kerridge, P. (1935) Speech defects and deafness. SPEECH (Whole issue, single issue).

Kerrison, P.D. (1918) The tests for malingering in defective hearing. LARYNGOSCOPE, 28, 662-665.

Kirk, R.E. (1968) EXPERIMENTAL DESIGN PROCEDURES FOR THE BEHAVIORAL SCIENCES, Brooks/Cole, Belmont, California.

Kitto, J. (1860) THE LOST SENSES, Oliphant, Edinburgh.

Klatt, D.H. (1973) Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 53, 8-16.

Klatt, D.H. (1979) Lexical representations for speech production. INTERNATIONAL SYMPOSIUM ON THE COGNITIVE REPRESENTATION OF SPEECH, EDINBURGH UNIVERSITY.

Klein, G.S. (1965) On hearing one's own voice. In Schur, M. (Ed.) DRIVES, AFFECTS, BEHAVIOR VOL 2, International University Press, New York.

Kryter, K.D (1970) THE EFFECTS OF NOISE ON MAN, Academic Press, London.

Lackner, J.R. (1974) Speech production: evidence for corollary-discharge stabilization of perceptual mechanisms. PERCEPTUAL AND MOTOR SKILLS, 39, 899-902.

Ladefoged, P. THREE AREAS OF EXPERIMENTAL PHONETICS, Oxford University Press, London.

Lane, H. (1976) THE WILD BOY OF AVEYRON, Harvard University Press, Cambridge, Massachusetts.

Lane, H., Catania, A.C., & Stevens, S.S. (1961) Voice level: autophonic scale, perceived loudness, and effects of side tone. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 33, 160-168.

Lane, H., Tranel, B., & Sisson, C. (1969) Regulation of voice communication by sensory dynamics. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 47, 618-624.

Lane, H., & Tranel, B. (1971) The Lombard sign and the role of hearing in speech. JOURNAL OF SPEECH AND HEARING RESEARCH, 14, 677-709.

Lane, H., & Van Teslaar, A.P. (1974) Voice spectrum and sidetone spectrum. In Moskowitz, H.R., Scharf, B., & Stevens, J.C. (Eds.) SENSATION AND PERCEPTION. PAPERS IN HONOR OF S.S. STEVENS, Reidel, Dordrecht.

Lashley, K.S. (1951) The problem of serial order in behavior. In Jeffress, L.A. (Ed.) CEREBRAL MECHANISMS IN BEHAVIOR. THE HIXON SYMPOSIUM. Wiley, New York and London.

Laver, J. (1976) Language and nonverbal communication. In Carterette, E.C., & Friedman, M.P. (Eds.) HANDBOOK OF PERCEPTION VOL. VII, LANGUAGE AND SPEECH, Academic Press, New York.

Lee, B.S. (1950a) Some effects of side-tone delay. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 22, 639-640.

Lee, B.S. (1950b) Effects of delayed speech feedback. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 22, 824-826.

Lee, B.S. (1951) Artificial stutter. JOURNAL OF SPEECH AND HEARING DISORDERS, 16, 53-55.

Lee, D.N. (1975) Vision in movement and balance. NEW SCIENTIST, 9th January, 1975.

Lefevre, E., Starck, R., Lambert, W.E., & Genesee, F. (1977) Lateral eye movements during verbal and nonverbal dichotic listening. PERCEPTUAL AND MOTOR SKILLS, 44, 1115-1122.

Lenneberg, E.H. (1964) Speech as a motor skill with special reference to nonaphasic disorders. In Pribram, K.H. (Ed.) BRAIN AND BEHAVIOUR VOL. 4: ADAPTATION, Penguin, Harmondsworth.

Lenneberg, E.H. (1967) BIOLOGICAL FOUNDATIONS OF LANGUAGE, Wiley, New York.

Lewes, G.H. (1860) THE PHYSIOLOGY OF COMMON LIFE, VOL. 2, Tauchnitz, Leipzig.

Li, C.C. (1964) INTRODUCTION TO EXPERIMENTAL STATISTICS, McGraw-Hill, New York.

Lightfoot, C., & Morrill, S.N. (1949) Loudness of speaking: the effect of the intensity of sidetone upon the intensity of the speaker. JOINT PROJECT REPORT NO. 4, NAVAL AIR STATION, U.S. NAVAL SCHOOL OF AVIATION MEDICINE AND RESEARCH, Pensacola, Florida.

Lombard, E. (1909) Letter to the French Academy of Sciences, 30th August, No. 7527.

Lombard, E. (1911) Le signe de l'élévation de la voix. ANNALES DES MALADIES DE L'OREILLE, DU LARYNX, DU NEZ, ET DU PHARYNX, 37, 101-109.

Lombard, E., & Baldenweck, L. (1916-17) Traitement de quelques cas de mutité hystéro-traumatique par le procédé de la suppression de contrôle auditif de la voix (méthode de Lombard). ARCHIVES DE MEDICINE ET PHARMACIE MILITAIRES, 66, 501-503.

Lower, M.C., & Martin, A.M. (1976) Effect of noise, visual cues and hearing protection on live speech communication. PROCEEDINGS OF THE INSTITUTE OF ACOUSTICS, Edinburgh University, September 1976, 2.9.1.-2.9.4.

Lucas Keene, M.F. (1957) Nerve endings in human laryngeal muscles. JOURNAL OF ANATOMY, 91, 590.

Luchsinger, R., & Arnold, G.E. (1965) VOICE. SPEECH. LANGUAGE, Constable, London.

MacKay, D.G. (1968) Metamorphosis of a critical interval: age-linked changes in the delay in auditory feedback producing maximal disruption with speech. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 43, 811-821.

MacKay, D.G. (1970) How does language familiarity influence stuttering under delayed auditory feedback? PERCEPTUAL AND MOTOR SKILLS, 30, 655-669.

Mackenzie, M. (1893) Speech and song. In ESSAYS, Sampson Low and Marston, London.

MacNeilage, P.F. (1972) Speech physiology. In Gilbert, J.H. (Ed.) SPEECH AND CORTICAL FUNCTIONING, Academic Press, New York..

MacNeilage, P.F., & Ladefoged, P. (1976) The production of speech and language. In Carterette,E.C.,& Friedman, M.P. (Eds.) HANDBOOK OF PERCEPTION VOL. VII, LANGUAGE AND SPEECH, Academic Press, New York.

Mahl, G.F. People talking when they can't hear their voices. In Siegman, A.W., & Pope, B. (Eds.) STUDIES IN DYADIC COMMUNICATION, Pergamon, New York.

Malloy, M.M. (1953) Investigation of the relationship between the pitch of side-tone and the frequency of the voice. M.A. thesis, Ohio State University.

Manning, W.H., Louko, L.J., & DiSalvo, V.S. (1978) A right-ear effect for auditory feedback control of children's newly acquired phonemes. JOURNAL OF SPEECH AND HEARING RESEARCH, 21, 580-588.

Matthews, L.H. (1952) Dampier: pirate and naturalist. NATURE, 170, 408-409.

McCall, G.N. (1973) Acoustic impedance measurement in the study of patients with spasmodic dysphonia. JOURNAL OF SPEECH AND HEARING DISORDERS, 38, 250-255.

McCall, G.N., & Rabuzzi, D.D. (1973) Reflex contraction of middle-ear muscles secondary to stimulation of laryngeal nerves. JOURNAL OF SPEECH AND HEARING RESEARCH, 16, 56-61.

McCroskey, R. (1958) The relative contributions of auditory and tactile cues to certain aspects of speech. SOUTHERN SPEECH JOURNAL, 24, 84-90.

McFarland, D.J. (1971) FEEDBACK MECHANISMS IN ANIMAL BEHAVIOUR. Academic Press, London and New York.

Melnick, W. (1965) The effect of stapedectomy on the loudness of one's own voice. JOURNAL OF SPEECH AND HEARING RESEARCH, 8, 223-234.

Mendelson, J., Kubzansky, P.E., Leiderman, P.H., Wexler, D., & Solomon, P. (1961) Physiological and psychological aspects of sensory deprivation - a case analysis. In Solomon, P.et al (Eds.) SENSORY DEPRIVATION, Harvard University Press, Cambridge, Massachusetts.

Miller, G.A., Galanter, E., & Pribram, K.H. (1960) PLANS AND THE STRUCTURE OF BEHAVIOR, Holt, Rinehart and Winston, New York.

Molinari, G.A., & Pivotti, G.S. (1963) The cochleorecurrential reflex. ARCHIVES OF OTOLARYNGOLOGY, 77, 92-100.

Moray, N. (1969) ATTENTION, Hutchinson, London.

Morton, J. (1970) A functional model for memory. In Norman, D.A. (Ed.) MODELS OF HUMAN MEMORY, Academic Press, London and New York.

Moscovici, S., & Plon, M. (1966) Les situations colloques. BULLETIN DE PSYCHOLOGIE, 19, 702-722.

Mott, F.W. (1910) THE BRAIN AND THE VOICE IN SPEECH AND SONG, Harper and Bros., London and New York.

Mowrer, O.H. (1958) Hearing and speaking: an analysis of language learning. JOURNAL OF SPEECH AND HEARING DISORDERS, 23, 143-152.

Myers, T.I., Murphy, D., & Terry, D.F. (1962) The role of expectancy in subjects' responses to sustained sensory deprivation. AMERICAN PSYCHOLOGICAL ASSOCIATION, St. Louis, September 1962.

Nottingham, J. (1857) DISEASES OF THE EAR, Churchill, London.

Paulhan, F. (1866) Le langage intérieur. REVUE PHILOSOPHIQUE, January 1866, p. 34.

Penn, J. (1955) Voice and speech patterns of the hard of hearing. ACTA OTOLARYNGOLOGICA, SUPPLEMENT 124.

Penzer, N.M. (1970) A NEW VOYAGE ROUND THE WORLD. BY WILLIAM DAMPIER, Da Capo Press, New York.

Peters, R.W. (1954) The effect of changes in side-tone delay and level upon rate of oral reading of normal speakers. JOURNAL OF SPEECH AND HEARING DISORDERS, 19, 483-491.

Peters, R.W. (1955) The effect of filtering of sidetone upon speaker intelligibility. JOURNAL OF SPEECH DISORDERS, 20, 371-375.

Peterson, G.E. (1954) Systematic research in experimental phonetics 4: The evaluation of speech signals. JOURNAL OF SPEECH AND HEARING DISORDERS, 19, 158-168.

Peterson, G.E. (1955) An oral communication model. LANGUAGE, 31, 414-427.

Piaget, J. (1948) LA NAISSANCE DE L'INTELLIGENCE CHEZ L'ENFANT, Delachoux and Niestlé, Paris.

Piaget, J. (1953) THE ORIGIN OF INTELLIGENCE IN THE CHILD, Routledge and Kegan Paul, London.

Pickett, J.M. (1956) Effects of vocal force on the intelligibility of speech sounds. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 28, 902-905.

Pickett, J.M. (1958) Limits of direct speech communication in noise. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 30, 278-281.

Pillsbury, W.B., & Meader, C.L. (1928) THE PSYCHOLOGY OF LANGUAGE, Appleton, London and New York.

Poulton, E.C. (1957) On prediction in skilled movements. PSYCHOLOGICAL BULLETIN, 54, 467-478.

Pritchard, U. (1886) DISEASES OF THE EAR, Lewis, London.

Rawnsley, A.I., & Harris, J.D. (1954) Comparative analysis of normal speech and speech with delayed sidetone by means of sound spectrograms. U.S. NAVY MEDICAL RESEARCH LABORATORY REPORT, PROGRAM NO. NM 003 041.56.03, NO. 248.

Ringel, R.L., & Steer, M.D. (1963) Some effects of tactile and auditory alterations on speech output. JOURNAL OF SPEECH AND HEARING RESEARCH, 6, 369-378.

Roehrig, W.C. (1965) Addition of controlled distortion to delay of auditory feedback. PERCEPTUAL AND MOTOR SKILLS, 21, 407-413.

Roosa, D.B.St.J (1873) TREATISE ON DISEASES OF THE EAR, William Wood, New York.

Rose, S. (1971) THE CHEMISTRY OF LIFE, Penguin, Harmondsworth.

Ross, A (Ed.) (1965) ROBINSON CRUSOE. BY DANIEL DEFOE, Penguin, Harmondsworth.

Rotman, B. (1978) A hymn to equilibrium. TIMES LITERARY SUPPLEMENT, 7th April, 1978, 373.

Rouse, R.O., & Tucker, G.R. (1966) An effect of delayed auditory feedback on speech in American and foreign students. JOURNAL OF SPEECH AND HEARING RESEARCH, 9, 456-460.

Rubenovitch, P., & Pastier, J. (1938) L'epreuve de Lombard appliquee en psychiatrie (contribution a l'etude des reflexes conditionn els). ANNALES DE MEDICINE PSYCHOLOGIQUE, 96, 116-121.

Ruch, T.C. (1951) Motor systems. In Stevens, S.S. (Ed.) HANDBOOK OF EXPERIMENTAL PSYCHOLOGY, Wiley, New York.

Rudolph, G. (1961) Spinal nerve endings (proprioceptors) in the human vocal muscle. NATURE, 190, 762.

Rutherford, D. (1967) Auditory motor learning and the acquisition of speech. AMERICAN JOURNAL OF PHYSICAL MEDICINE, 46, 245-251.

Salter, D. (1973) Capacity and delayed speech. BRITISH JOURNAL OF PSYCHOLOGY, 64, 193-198.

Scripture, E.W. (1902) THE ELEMENTS OF EXPERIMENTAL PHONETICS, Charles Scribner's Sons, New York; Edward Arnold, London.

Shane, M.S. (1955) Effect on stuttering of alteration in auditory feedback. In Johnson, W.M. (Ed.) STUTTERING IN CHILDREN AND ADULTS, University of Minnesota Press, Minneapolis.

Shearer, W.M. (1966) Speech: behavior of middle ear muscle during stuttering. SCIENCE , 152, 1280.

Shearer, W.M., & Simmons, F.B. (1965) Middle ear activity during speech in normal speakers and stutterers. JOURNAL OF SPEECH AND HEARING RESEARCH, 8, 203-207.

Sherrard, C.A. (1975) Stuttering as false alarm responding. BRITISH JOURNAL OF DISORDERS OF COMMUNICATION, 10, 83-91.

Sherrington, C.S., & Mott, F.W. (1895) Experiments upon the influence of sensory nerves upon movement and nutrition of the limbs. PROCEEDINGS OF THE ROYAL SOCIETY, 57, 481-488.

Shipman, J.C. (1962) WILLIAM DAMPIER SEAMAN-SCIENTIST. University of Kansas Libraries, Lawrence, Kansas.

Siegel, G.M., & Pick, H.L. (1974) Auditory feedback in the regulation of voice. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 56, 1618-1624.

Silverman, F.H., & Goodban, M.T. (1972) The effect of auditory masking on the fluency of normal speakers. JOURNAL OF SPEECH AND HEARING RESEARCH, 15, 543-546.

Sluckin, W. (1954) MINDS AND MACHINES, Penguin, Harmondsworth.

Smith, S., & Guthrie, E.R. (1921) GENERAL PSYCHOLOGY IN TERMS OF BEHAVIOR, Appleton, New York.

Stanton, J.B. (1958) The effects of delayed auditory feedback on the speech of aphasic patients. SCOTTISH MEDICAL JOURNAL, 3, 378.

Stark, L. (1968) NEUROLOGICAL CONTROL SYSTEMS. STUDIES IN BIOENGINEERING, Plenum Press, New York.

Stevenson, R.S., & Guthrie, D.J. (1949) A HISTORY OF OTO-LARYNGOLOGY, Livingstone, Edinburgh.

Stratton, G.M. (1896) Some preliminary experiments in vision without inversion of the retinal image. PSYCHOLOGICAL REVIEW, 3, 611-617.

Strattón, G.M. (1897) Vision without inversion of the retinal image. PSYCHOLOGICAL REVIEW, 4, 341-360 and 463-481.

Sutton, S., & Chase, R.A. (1961) White noise and stuttering. JOURNAL OF SPEECH AND HEARING RESEARCH, 4, 72.

Swinnerton, F. (1956) BACKGROUND WITH CHORUS, Hutchinson, London.

Taub, E., & Berman, A.J. (1968) Movement and learning in the absence of sensory feedback. In Freedman, S.J. (Ed.) THE NEUROPSYCHOLOGY OF SPATIALLY ORIENTED BEHAVIOR, Dorsey Press, Homewood, Illinois.

Tournier, M. (1974) FRIDAY, OR THE OTHER ISLAND, Penguin, Harmondsworth.

Toynbee, J. (1860) DISEASES OF THE EAR, Churchill, London.

Treisman, A.M. (1965) Reading rate, word information and auditory monitoring of speech. NATURE, 205, 1297-1300.

Trudgill, P. (1974) SOCIOLINGUISTICS, Penguin, Harmondsworth.

Van Galen, G.P., & Ten Hoopen, G. (1976) Speech control and single channelness. ACTA PSYCHOLOGICA, 40, 245-255.

Vartanian, A. (1960) LA METTRIE'S 'L'HOMME MACHINE', A STUDY IN THE ORIGINS OF AN IDEA, Princeton University Press, Princeton, New Jersey.

Vygotsky, L.S. (1962) THOUGHT AND LANGUAGE, MIT Press, Cambridge, Massachusetts.

Waugh, E. (1948) WHEN THE GOING WAS GOOD, Duckworth, London.

Webster, R.L., & Dorman, M.F. (1970) Decreases in stuttering frequency as a function of continuous and contingent forms of auditory masking. JOURNAL OF SPEECH AND HEARING RESEARCH, 13, 82-86.

Wechsler, D. (1955) MANUAL FOR THE WECHSLER ADULT INTELLIGENCE SCALE, The Psychological Corporation, New York.

Welford, A.T. (1968) FUNDAMENTALS OF SKILL, Methuen, London.

Wiener, N. (1948) CYBERNETICS; OR CONTROL AND COMMUNICATION IN THE ANIMAL AND THE MACHINE, Wiley, New York.

Wiener, N. (1950) THE HUMAN USE OF HUMAN BEINGS, Eyre and Spottiswoode, London.

Williams, C.E., & Hecker, M.H.L. (1968) Relation between intelligibility scores for four test methods and three types of speech distortion. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 44, 1002-1006.

Woodworth, R.S. (1899) The accuracy of voluntary movement. PSYCHOLOGICAL MONOGRAPHS, 3, NO. 2.

Zubek, J.P., & MacNeill, M. (1967) Perceptual deprivation phenomena: role of the recumbent position. JOURNAL OF ABNORMAL PSYCHOLOGY, 72, 147-150.

Zwislocki, J., (1953) Acoustic Attenuation Between the Ears. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, 25, 152-9